

THE DIAGNOSTIC THINKING PROCESS IN MEDICAL
EDUCATION AND CLINICAL PRACTICE

A study of medical students, house officers
and registrars with special reference to
endocrinology and neurology

by

Janet Gale

Presented in partial fulfilment of the requirements
for the Ph.D degree. London University Institute
of Education, 1980.



For Rod

ABSTRACT

THE DIAGNOSTIC THINKING PROCESS IN MEDICAL EDUCATION AND CLINICAL PRACTICE. A study of medical students, house officers and registrars with special reference to endocrinology and neurology.

by

Janet Gale

Aims

1. To add clarity, specificity and breadth to current descriptions of the diagnostic thinking process as hypothesis generation and testing.
2. To propose possible pedagogical strategies for efficient and effective development of appropriate diagnostic thinking processes.

Subjects and Methods

The study was in two complementary parts:

1. Quantitative data. Structured questionnaires in endocrinology and neurology allowed analysis of the relative contributions of (i) factual knowledge, (ii) interpretation of symptoms and signs, and (iii) selecting and testing diagnostic possibilities, towards predicting diagnostic ability in 35 final year clinical medical students and 35 registrars per questionnaire.
2. Qualitative data. Introspective account gathering by videotape stimulated recall of clinical interviews yielded data concerning the diagnostic thinking processes of 22 final year medical students, 22 house officers and 22 registrars.

The research design allowed analysis of the development of the diagnostic thinking process, the relative effects of medical education and clinical practice and comparison of these in two specialities.

Results

After parametric, non-parametric and content analyses, the results of the two studies were related to yield a unified explanation and description of the diagnostic thinking process in medical education and clinical practice. Its fundamental psychological features are identified as structure and extrapolation. Three stages are identified and discussed in terms of wider psychological theory. Speciality specific and generalised cognitive processes are identified. Development of the diagnostic thinking process is described in terms of increasing equilibration of the skills measured in the quantitative study. The qualitative study shows that students, house officers and registrars are in command of the same range of cognitive processes although their relative contributions and associated appropriateness and accessibility of content might vary between groups yielding different degrees of diagnostic accuracy. The diagnostic thinking process in all groups of subjects is shown to be considerably more complex than previous descriptions have suggested. Pedagogical implications are discussed.

ACKNOWLEDGMENTS

My grateful thanks go to:

Emeritus Professor W.D. Wall for his invaluable guidance and encouragement through every stage of the work, and for being exactly the supervisor I needed.

Those who inspired my original interest or communicated with me while the work was in progress, especially Professor Howard Barrows for motivational discussions, Dr. Vic Neufeld and Dr. Geoff Norman. The thanks extend to Professor A.S. Elstein for his published work which continues to be a source of stimulation and object of my *jalousie de métier*.

Dr. Philip Marsden without whose specialist knowledge, clinical experience, hard work and good temper, there would have been little data, less accuracy, no pre-diagnostic or diagnostic categories and possibly no thesis. As both colleague and friend, I owe him unredeemable debts.

Dr. Leslie Sedal whose work in writing the neurology questionnaire was formative for the quantitative part of the study.

All those who gave assistance in data collection, especially Professor Sir John Walton, Professor David Shaw and Dr. John Anderson all of the University of Newcastle Upon Tyne Medical School, Dr. M.J.G. Harrison of the Middlesex Hospital School of Medicine, and the many clinicians and medical teachers who have distributed and collected questionnaires within their own departments.

All those who donated their technical expertise during account gathering sessions, especially Mr. Tom Harlow of King's College Hospital Medical School and Mr. Ray Leng, Steve and Girish of the Department of Medical Photography, Greenwich District Hospital.

Without them, there would have been no qualitative study.

All the students, house officers and registrars who agreed to be subjects of the study. I hope that they, too, gained something from their involvement.

Sue Maton who spent many tedious hours marking, timing and transcribing and became Sue Randall during the task. The raw data owes its identity to her.

Manjula Varsani who walked miles around my office collating questionnaires and my husband who learned to make electrostencils and run duplicating machines for them.

The BLAT Centre for Health and Medical Education for their tolerance and postage.

Jackie Freeman for her painstaking calligraphy.

My friend, Kathy Kilcoyne, for her great skill and patience in turning my manuscript into beautifully organised typed pages, and to her sister, Marion, for assisting in that task. All my work is mediated through them. It is they who present it to you.

All those who never seemed to lose faith that this thesis would eventually be finished.

Finally, my husband, who has cooked, shopped, cleaned, laundered, gardened, spent successive solitary evenings and weekends, had no holiday and still brought me coffee and cake for more than a year so that I could work on and finish this. The preparation of his thesis he did not inflict on me; mine has disrupted his life. It is dedicated to him.

With us ther was a DOCTOR of PHISIK;
In all this world ne was ther noon him lik,
To speke of phisik and of surgerye;
For he was groundes in astronomye.
He kepte his pacient a ful greet deel
In houres, by his magyk natureel.
Wel koude he fortunen the ascendent
Of his ymages for his pacient.
He knew the cause of everich maladye,
Were it of hoot, or cold, or moyste or drye,
And where they engendred and of what humour;
He was a verray parfit practisour.

Chaucer

TABLE OF CONTENTS

	Page
<u>CHAPTER ONE. Introduction to the Problem</u>	
1.1 Reasons for the Research	24
1.2 Aims and Outline of the Research	28
1.3 Areas Excluded from the Research	29
1.4 Content and Organisation of the Thesis	31
1.5 Summary	32
 <u>CHAPTER TWO. Review of the Literature 1. Mathematical and Logical Models</u>	
2.1 Statistical Models	33
2.1.1 Linear Models	34
2.1.2 Non-linear Models	39
2.2 Other Models	41
2.3 Statistical versus Clinical Prediction	43
2.4 Summary	46
Figure 2.1 Simplified Representation of the Lens Model	40
Figure 2.2 Illustration of Clinicians' and Computer's Pathways to Diagnosis, Plotted Within a Triangle (after Taylor et al, 1971)	44
 <u>CHAPTER THREE. Review of the Literature 2. Cognitive Studies</u>	
3.1 Hypothesis Generation and Testing	48
3.1.1 The Michigan Medical Inquiry Project	48
3.1.2 The McMaster Studies	56
3.2 Pattern Recognition	61
3.3 Confirming, Refuting and Irrelevant Information	64
3.4 Other Factors	67

	Page
3.5 Summary	70
3.6 Implications for the Present Research	74
 <u>CHAPTER FOUR. Review of the Literature 3. Teaching the Diagnostic Thinking Process</u>	
4.1 Theoretical Aspects of Teaching and Learning	76
4.1.1 Structure in Learning	77
4.1.2 Transfer of Learning	81
4.1.3 Problem Solving and Learning	86
4.1.4 Conclusions	89
4.2 Approaches in Teaching the Diagnostic Thinking Process	91
4.2.1 The Integrated Curriculum	92
4.2.2 Problem Based Learning	97
4.2.3 Training in Hypothesis Generation	106
4.2.4 Cognitive Skills Training	110
4.2.5 Computer Assisted Training	111
4.3 Summary and Conclusions.	116
 <u>CHAPTER FIVE. Research Design: Hypotheses, Research Methods, Subjects, Design Validity</u>	
5.1 Hypotheses	119
5.2 Development of the Current Approach	123
5.2.1 Introduction	124
5.2.2 Clinical Problems in Propositional Logic	124
5.2.3 Development of the Current Approach	128
5.3 Questionnaires in Endocrinology and Neurology - Rationale and Description	129
5.3.1 Rationale	130
5.3.2 Description of Format and Structure	132

	Page
5.4 Accounts by Videotape Stimulated Recall	132
5.4.1 Rationale	132
5.4.2 Description	134
5.5 Relationship Between the Questionnaire Study and the Account Gathering Study	134
5.6 Subjects	135
5.6.1 The Questionnaire Study	135
5.6.2 The Account Gathering Study	138
5.6.3 Comparability of Questionnaire and Account Gathering Subjects	139
5.7 Design Validity	140
5.7.1 Internal Validity	140
5.7.2 External Validity	142
5.8 The Acceptable Level of Significance: $p < .01$	143
5.9 Summary and Conclusions	144
Table 5.1 Questionnaires in Endocrinology and Neurology: Age and Sex of Subjects	136
Table 5.2 Subjects of the Complementary Studies	139
<u>CHAPTER SIX. The Questionnaire Study 1. Validity, Reliability and Data Analysis</u>	
6.1 Validity of the Questionnaires	148
6.1.1 Criterion Oriented Validity	148
6.1.2 Content Validity	150
6.1.3 Construct Validity	152
6.1.4 Validity by Assumption	152
6.2 Validation by Test and Item Analysis	152
6.2.1 Test Difficulty per Group	154
6.2.2 The Phi Coefficient (ϕ)	155
6.3 Reliability and Homogeneity of the Questionnaires	156

	Page
6.4 Conclusions	158
6.5 Statistical Analysis of Data	158
6.5.1 Stepwise Linear Regression	160
6.5.2 Two-way Analysis of Variance	161
6.5.3 Table of Differences Between Groups in Mean Scores per Section with t-tests. Supported by Discriminant Analysis	162
6.5.4 One-way Analysis of Variance with Scheffé Test Comparing Scores on Sections for Students and Registrars Separately	163
6.6 Summary	163
Figure 6.1 Statistical Tests and Associated Hypotheses for Questionnaires in Endocrinology and Neurology Separately	160
Table 6.1 Questionnaire in Endocrinology - Topics, Number of Questions, Range of Raw Scores Possible	146
Table 6.2 Questionnaire in Neurology - Topics, Number of Questions, Range of Raw Scores Possible	147
Table 6.3 Intraclass Correlation Coefficient (R). Summary of Results for Endocrinology Questionnaire	150
Table 6.4 Intraclass Correlation Coefficient (R). Summary of Results for Neurology Questionnaire	150
 <u>CHAPTER SEVEN. The Questionnaire Study 2. Results</u>	
7.1 Stepwise Linear Regression	165
7.1.1 Endocrinology Students	167
7.1.2 Endocrinology Registrars	169
7.1.3 Neurology Students	170

	Page
7.1.4 Neurology Registrars	171
7.1.5 Comparison of Groups	172
7.2 Two-way Analysis of Variance	174
7.2.1 Endocrinology	174
7.2.2 Neurology	175
7.3 One-way Analysis of Variance with Table of Observed Differences in Mean Scores, Comparing Students' and Registrars' Scores on Each Section Separately	176
7.3.1 Endocrinology	177
7.3.2 Neurology	178
7.3.3 Comparison of Endocrinology and Neurology	178
7.4 One-way Analysis of Variance with Scheffé Test, Comparing Scores on Sections for Students and Registrars Separately	180
7.4.1 Endocrinology Students	180
7.4.2 Endocrinology Registrars	182
7.4.3 Neurology Students	183
7.4.4 Neurology Registrars	184
7.4.5 Comparison of Groups	185
7.5 Summary	186
Figure 7.1 Diagrammatic Representation of Relative Contributions of Sections A, B and C in Predicting Scores on Section D (Free Stepwise Regression)	173
Table 7.1 Correlation Matrix Between all Variables - Endocrinology Students	165
Table 7.2 Correlation Matrix Between all Variables - Endocrinology Registrars	166
Table 7.3 Correlation Matrix Between all Variables - Neurology Students	166

	Page
Table 7.4 Correlation Matrix Between all Variables - Neurology Registrars	166
Table 7.5 Relative Incidence of Statistically Significant Correlation Coefficients Across Groups of Subjects	167
Table 7.6 Summary of Free Stepwise Regression Analysis - Endocrinology Students	168
Table 7.7 Values of R and R^2 for Each Predictor Variable Independently of all Others - Endocrinology Students	168
Table 7.8 Summary of Free Stepwise Regression Analysis - Endocrinology Registrars	169
Table 7.9 Summary of Free Stepwise Regression Analysis - Neurology Students	170
Table 7.10 Values of R and R^2 for Each Predictor Variable Independently of all Others - Neurology Students	171
Table 7.11 Summary of Free Stepwise Regression Analysis - Neurology Registrars	171
Table 7.12 Values of R and R^2 for Each Predictor Variable Independently of all Others - Neurology Registrars	172
Table 7.13 Two-way Analysis of Variance Summary Table - Endocrinology	175
Table 7.14 Two-way Analysis of Variance Summary Table - Neurology	176
Table 7.15 Observed Differences in Mean Scores per Section Between Students and Registrars - Endocrinology	177
Table 7.16 Observed Differences in Mean Scores per Section Between Students and Registrars - Neurology	178
Table 7.17 Relative Magnitude and Ratios of Observed Differences Between Students and Registrars on Mean Scores per Section for Endocrinology and Neurology	179

	Page
Table 7.18 One-way Analysis of Variance Summary Table for Scores on Sections - Endocrinology Students	181
Table 7.19 Results of Scheffé Method of Complete Set of Comparisons - Endocrinology Students	181
Table 7.20 One-way Analysis of Variance Summary Table for Scores on Sections - Endocrinology Registrars	182
Table 7.21 Results of Scheffé Method of Complete Set of Comparisons - Endocrinology Registrars	182
Table 7.22 One-way Analysis of Variance Summary Table for Scores on Sections - Neurology Students	183
Table 7.23 Results of Scheffé Method of Complete Set of Comparisons - Neurology Students	183
Table 7.24 One-way Analysis of Variance Summary Table for Scores on Sections - Neurology Registrars	184
Table 7.25 Results of Scheffé Method of Complete Set of Comparisons - Neurology Registrars	184
 <u>CHAPTER EIGHT. The Questionnaire Study 3. Discussion of Results</u>	
8.1 Overview of Issues for Discussion	187
8.2 Introduction to the Discussion	188
8.3 Hypothesis 1 and Hypothesis 5	190
8.3.1 Endocrinology	191
8.3.2 Neurology	196
8.3.3 Comparison of Endocrinology and Neurology	197
8.4 Hypothesis 2	199
8.4.1 Endocrinology	200
8.4.2 Neurology	200
8.4.3 Comparison of Endocrinology and Neurology	201

	Page
8.5 Hypothesis 3	202
8.5.1 Endocrinology	203
8.5.2 Neurology	203
8.5.3 Discussion of Endocrinology and Neurology Results	204
8.6 Hypotheses 4 and 6	204
8.6.1 Endocrinology	205
8.6.2 Neurology	206
8.6.3 Comparison of Endocrinology and Neurology	207
8.7 Hypothesis 7	207
8.8 Hypothesis 8	208
8.9 Hypothesis 9	212
8.10 Summary and Conclusions	215
Figure 8.1 Diagrammatic Representation of the Assumed Relationships Between Scores on Sections A,B, C and D of the Multiple Choice Questionnaires	192
Table 8.1 The Relationship Between Students' and Registrars' Skills Structures, Cognitive Processes and Diagnostic Acumen (in relative, not absolute, terms)	201
 <u>CHAPTER NINE. The Account Gathering Study 1. Description, Validity, Reliability and Data Analysis</u>	
9.1 Description and Discussion of the Implementation of the Method	219
9.2 Validity of the Method	225
9.2.1 The Method of Introspection	226
9.2.2 The Method of Stimulated Recall	234
9.2.3 Selection of Content (Patients)	237
9.2.4 Other Methodological Aspects: Videorecording and Account Gathering	239

	Page
9.2.5 Omission of the Physical Examination	241
9.2.6 Conclusions	242
9.3 Data Analysis	242
9.3.1 The χ^2 Test for k Independent Samples	244
9.3.2 The Kruskal-Wallis One-Way Analysis of Variance by Ranks	245
9.4 Summary	245
Table 9.1 Accounts by Stimulated Recall. Length (in Seconds) of Patient Interviews for Students, House Officers and Registrars	223
Table 9.2 Accounts by Stimulated Recall. Length (in Seconds) of Replay Periods for Students, House Officers and Registrars	223
Table 9.3 Accounts by Stimulated Recall. Number of Stops Made for Account Gathering for Students, House Officers and Registrars	223
Table 9.4 Accounts by Stimulated Recall. Total Number of Utterances Made During Each Replay Period for Students, House Officers and Registrars	224
Table 9.5 Subjects' Estimates of the Relationship of Video- recorded Interviews to Usual Performance (22 subjects per group)	240
Table 9.6 Hypotheses and Categories Subjected to the χ^2 Test for k Independent Samples and the Kruskal-Wallis One-way Analysis of Variance, Respectively	244

	Page
<u>CHAPTER TEN. The Account Gathering Study 2. Results</u>	
10.1 Content Analysis: Categories and Indicators	247
10.1.1 Hypothesis 10 Categories and Indicators	247
10.1.2 Hypothesis 11 Categories and Indicators	249
10.1.3 Hypothesis 12 Categories and Indicators	249
10.1.4 Hypothesis 13 Categories and Indicators	250
10.1.5 Hypothesis 14 Categories and Indicators	251
10.1.6 Hypothesis 15 Categories and Indicators	253
10.1.7 Hypothesis 16 Categories and Indicators	255
10.2 Statistical Analyses	256
10.2.1 Hypothesis 10. Parametric Two-Way Analysis of Variance	256
10.2.2 Hypothesis 11. Parametric One-Way Analysis of Variance and χ^2	258
10.2.3 Hypothesis 12. χ^2	262
10.2.4 Hypothesis 13. Kruskal-Wallis	262
10.2.5 Hypothesis 14. χ^2 and Kruskal-Wallis	264
10.2.6 Hypothesis 15. Kruskal-Wallis	267
10.2.7 Hypothesis 16. Kruskal-Wallis	268
10.3 Summary	270
Table 10.1 Two-Way Analysis of Variance Summary Table. Hypothesis 10, Categories (a) and (b)	257
Table 10.2 Summary of Raw Results for Categories 10(a) and 10(b)	257
Table 10.3 Length (in seconds) of the First Replay Period	259
Table 10.4 One-Way Analysis of Variance Summary Table for Length of First Replay Period	259

	Page
Table 10.5 Observed Frequency of Response in Categories 11(a), (b) and (c) for Each Group	260
Table 10.6 Frequency of Single and Multiple Responses to Information Elicited During the First Replay Period, Categories 11(a), (b) and (c)	261
Table 10.7 Frequencies of Subjects' Responses in Single and Combined Categories 11(a), (b) and (c)	261
Table 10.8 Summary of χ^2 test to Determine the Independence of the Variables: Status and Response Category for 11(a), (b) and (c)	262
Table 10.9 Summary of χ^2 test to Determine the Independence of Variables: Status and Frequency of Response per Category for 12(a), (b) and (c)	263
Table 10.10 Frequency of Responses per Group in Categories 13(a) and 13(b) Separately and Conjointly and Number of Subjects Making Responses in Neither Category	263
Table 10.11 Summary of χ^2 test to Determine the Independence of the Variables: Status and Strategies of Use of Categories 14(a), (b) and (c)	265
Table 10.12 Frequency of Subjects' Responses in Single and Combined Categories 14(a), (b) and (c)	266
Table 10.13 Observed Frequency of Use of Categories 15(a), (b) and (c)	268
Table 10.14 Frequency of Observed Responses in Categories 16(a) and (b) for all Groups	269
Table 10.15 Comparison of Observed Frequencies in Categories 16(a) and (b) with Arbitrary Theoretical Frequency of 14(c. 64 per cent)	269

CHAPTER ELEVEN. The Account Gathering Study 3. Discussion
of Results

11.1 Overview of the Issues for Discussion	271
11.2 Hypothesis 10	271
11.3 Hypothesis 11	277
11.4 Hypothesis 12	282
11.5 Hypotheses 13 and 17	285
11.6 Hypothesis 14	294
11.7 Hypothesis 15	299
11.8 Hypothesis 16	305
11.9 Hypothesis 18	307
11.9.1 Set	308
11.9.2 Failure to Make the Correct Interpretation	312
11.9.3 Designation of Irrelevance	316
11.9.4 Conclusions	318
11.10 Hypothesis 19	319
11.11 Hypothesis 20	320
11.12 Conclusions: Interpretative Value of the Present Study for Current Descriptions of the Diagnostic Thinking Process	322
Figure 11.1 Example of Reinterpretation (Cognitive Restructuring) of a Symptom (Diarrhoea) in the Presence of New Information. Category: 13(b). Subject: S.02	289
Figure 11.2 Example of Cognitive Restructuring of a Symptom (Nosebleed) by Embedding in a New Context With Complex Restructuring of an Information Cluster. Category: 13(b). Subject: S.04	290

- Figure 11.3 Example of Cognitive Restructuring From Cardiac
to Respiratory Context. Category: 13(b).
Subject: S.12 291
- Figure 11.4 Example of Initial Failure to Find the Appropriate
Extrapolated Context, Followed by Rejection of
Inappropriate Interpretation and Final Restructuring
after New Information Facilitates Correct Interpret-
ation. Diagnosis: Stokes Adams Attacks Due to
Ischaemic Heart Disease. Category:13(b). Subject:S.19 291
- Figure 11.5 Subjects' Comments on Their Use of the Routine Clinical
History 303
- Figure 11.6 Example of Set for Cardiovascular Disease after
Diagnosing Myocardial Infarct with Consequent Failure
to Pursue Symptoms in Other Systems or Identify Them
as not Cardiovascular. Actual Diagnoses: Myocardial
Infarct, Duodenal Ulcer, Cervical Spondylosis.
Subject: S.06 309
- Figure 11.7 Example of Set for Bladder Tumour (Carcinoma)
Causing Failure to Accept Significant Information
(Pernicious Anaemia) and Make Relevant Inquiry (Heart
Failure). Actual Diagnoses: Recurrent Urinary Tract
Infection with Calculi, Mild Heart Failure Secondary
to Pernicious Aneamia. Subject: HO.21 309
- Figure 11.8 Example of Set for Incorrect Diagnosis (Functional
Diarrhoea) Causing Misinterpretation and Ignoring
of Significant Information. Actual Diagnosis: Crohn's
Disease. Subject: R.14 311

Figure 11.9	Example of Failure to Extrapolate to the Correct Context (Asthma). Actual Diagnoses: Intrinsic (late onset) Asthma, Acute Bronchitis and Allergic Aspergellosis. Subject: R.09	313
-------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----

Figure 11.10	Example of Failure to Extrapolate to Correct Context. Actual Diagnosis: Pulmonary Tuberculosis with Hypoalbuminaemic Oedema. Subject: R.21	315
--------------	--------------------------------------------------------------------------------------------------------------------------------------------	-----

CHAPTER TWELVE. The Diagnostic Thinking Process: A Psychological Perspective Relating the Results of the Two Parallel Studies

12.1	Fundamental Psychological Features: Structure and Extrapolation	336
12.1.1	Structure	336
12.1.2	Extrapolation	357
12.2	Stages of the Diagnostic Thinking Process - A Psychological Perspective	354
12.2.1	Stage One: Initiation of Interpretations	355
12.2.2	Stage Two: Progress of the Diagnostic Thinking Process	367
12.2.3	Stage Three: Resolution	381
12.2.4	The Central Characteristics of the Three Stages of the Diagnostic Thinking Process	387
12.3	Developmental and Comparative Aspects of the Diagnostic Thinking Process	392
12.4	Summary and Conclusions	403

CHAPTER THIRTEEN. Implications of the Current Study for
Pedagogical Approaches in the Development
of the Diagnostic Thinking Process

13.1	Medical Education and Clinical Practice	405
13.1.1	The Current Relationship	405
13.1.2	The Potential Relationship	414
13.2	Aims in the Design of Teaching Strategies	414
13.3	Pedagogical Approaches	425
13.3.1	Contextual Considerations	425
13.3.2	Curriculum Content and the Role of the Teacher	436
13.3.3	The Pedagogy of Structure	439
13.3.4	The Pedagogy of Process and its Relationship to the Pedagogy of Structure	449
13.3.5	An Evaluation of the Proposed Pedagogical Approaches	460
13.4	Summary and Conclusions	465

CHAPTER FOURTEEN. Limitations of the Present Research and
Indications for Further Study

14.1	Limitations of the Present Research	467
14.2	Indications for Further Study	470
14.3	Summary	471

EPILOGUE 472

APPENDICES

Appendix 1	Information Given About the Features of Syncope and Epilepsy	473
Appendix 2	Questionnaire in Endocrinology - Content	488
Appendix 3	Questionnaire in Neurology - Content	489

	Page
Appendix 4 Rationale for Selection of Five Option Independent True False Multiple Choice Questions	505
Appendix 5 Policy on Guessing, Instructions to Respondents and Scoring Schedule for Questionnaires	507
Appendix 6 The Intraclass Correlation Coefficient: Nature, Purpose and Formula	511
Appendix 7 Calculation of the Intraclass Correlation Coefficient (R) for Endocrinology and Neurology Questionnaires Separately	513
Appendix 8 Calculation of the Test Difficulty Index	515
Appendix 9 Calculation of the Phi Coefficient	516
Appendix 10 Values of the Phi Coefficient (ϕ) for the Questionnaires in Endocrinology and Neurology	517
Appendix 11 Split Half Test of Reliability: Method and Correction Factor	519
Appendix 12 Notes on the Interpretation of Regression Summary Tables	520
Appendix 13 Discriminant Analysis as Applied to the Questionnaires in Endocrinology and Neurology	521
Appendix 14 The Scheffé Method as Applied to Data from the Questionnaires in Endocrinology and Neurology, after One-way Analysis of Variance	524
Appendix 15 Procedures for Normalisation of Raw Scores	525
Appendix 16 Discriminant Analysis Results and Discussion	526
Appendix 17 Example of Experimenter's Notes Taken During Simultaneous Viewing of a Patient-House Officer Encounter	536

	Page
Appendix 18 Transcript of an Account Gathering Session with a Student, with Experimenter's Notes	538
Appendix 19 Transcript of an Account Gathering Session with a House Officer, with Experimenter's Notes	544
Appendix 20 Transcript of an Account Gathering Session with a Registrar, with Experimenter's Notes	554
Appendix 21 Diagnoses Given in Each Patient's Medical Record - Cases Seen by Students	560
Appendix 22 Diagnoses Given in Each Patient's Medical Record - Cases Seen by House Officers	562
Appendix 23 Diagnoses Given in Each Patient's Medical Record - Cases Seen by Registrars	564
Appendix 24 Statistical Analyses of Patients' Diagnoses and Diagnoses Made Across Groups of Subjects	566
Appendix 25 Preliminary Considerations in the Interpretation of Protocols	568
Appendix 26 The Method of Content Analysis	570
Appendix 27 Establishment of Reliability and Validity of Content Analysis Categories and Raters	574
Appendix 28 Items with Designated Categories Selected for Reliability Study of Pre-diagnostic (P-d) and Diagnostic (D) Interpretations	578
Appendix 29 Content Analysis Raw Data: Observed Frequencies per Category per Group	582
Appendix 30 Raw Data of Content Analysis Using Categories 10(a) (pre-diagnostic interpretations) and 10(b) (diagnostic interpretations), Recording the Exact Number of Each Type per Subject per Group	583
<u>REFERENCES</u>	584/599

CHAPTER ONE

Introduction to the Problem

"Any heretic who talked to a Curriculum Committee about the skills that were needed to practise medicine, was accused of wanting to produce technicians ... not the medical scientists that they wanted their students to become. This produced the tacit and naive assumption that once the student's head was stuffed full of all the knowledge of basic and clinical medicine, and all the rituals of history taking and physical examination technique were well memorised, the mystique or art of the good physician would magically happen, as if by the touch of a fairy's wand". (Barrows, 1976).

The naive assumption, described here by a Professor of Neurology, is no longer made without challenge. This chapter presents an introduction and guide to the present study which seeks to evaluate and deepen our current understanding of the "art of the good physician" and its development in medical education and clinical practice.

Firstly, an explanation is offered of why the present study is a necessary pre-requisite for pedagogical development in relation to the diagnostic thinking process and clinical problem solving. Secondly, the aims of the research and an outline of its methodology are described. Thirdly, a brief statement is made of areas excluded from the study. Finally, the content and organisation of the thesis is set out.

1.1 Reasons for the Research

There is a well recognised and documented need to make the content and structure of medical education more appropriate to the student's future needs as a practising clinician (see sections 4.2 and 13.1.2). In particular, it is considered that diagnostic thinking processes (or clinical problem solving skills) should be dealt with more conscientiously by undergraduate medical education:

"There is no more important field in medicine than diagnosis. Without it, we are charlatans or witch doctors treating in the dark with potions and prayers. Yet there is no field more difficult to teach. Strange that this art and science has not attracted innumerable

theorists to make it more teachable!
Thousands are studying membrane transfer,
yet few strive to make a science of
diagnosis". (Cutler, 1979)

Here are posed both sides of the question. The skill of diagnosis must be taught, and our assumptions underlying such teaching must be sound (Berner et al, 1977). It is, therefore, necessary to define the nature of the cognitive process involved. The first major reason for the present research, then, can be stated quite simply. It is that if a subject is to be taught, efficiently and effectively, it must be understood by the teacher. Therefore, if the acquisition of effective diagnostic thinking processes and clinical problem solving skills is to be facilitated by the pedagogy of the undergraduate medical curriculum, those processes and skills must themselves be understood. The need to consider the cognitive skills of clinical practice and to provide a suitable learning environment for their acquisition or development has for long been recognised (Ellis, 1960). Shulman and Elstein (1975) cite Hammond (1971) on this point:

"... the teacher is frequently unaware of the real system he uses to make his expert judgments. He may even believe that he operates in a very different fashion from the way he actually does. Imagine the frustration of students who must learn to ignore what he says they should do and instead must infer the model of his judgments. Alas, claims Hammond, this is far more frequent in the teaching of clinical judgment than has been admitted or recognised". (p. 28)

Cutler (1979) echoes this opinion. Dudley (1969) considers that the clinical teacher needs to recognise the diagnostic process apart from the technique of collecting information and distinct from the imparting of factual knowledge. McWhinney (1972) also believes that students should be taught some theory of diagnosis so that they may engage in effective problem solving and decision making. While accepting these judgments, Iansek and Balla (1979) bring us back sharply to the reason for the present study. Balla, himself a neurologist, speaks from a particular vantage point:

"The majority of doctors are well aware of the many diagnostic decisions they make in their daily practice. However, not many doctors are aware of the underlying mechanisms involved in the making of such decisions, except to attribute such an ability to 'experience', 'art' or encyclopaedic knowledge. This nebulous attitude to their diagnostic ability is, of course, extended to an important facet of a doctor's responsibility: medical education. In this context, an incongruous situation is apparent in that the teacher is responsible for imparting to his students knowledge which the teacher does not fully understand or of which the teacher is unaware. Such an approach is obviously unacceptable, and requires rectification".

Such criticism of medical education is echoed by many medical teachers who point to the failure to teach clinical problem solving or reasoning skills (see the introduction to section 4.2 below). The need to avoid the incongruous situation described by Iansek and Balla (1979) is mentioned by workers whose proper field is the study of education (Gage, 1963):

"Any scheme of tuition or graded experience calculated to produce better insights into the act of judging must proceed from what we have discovered about the stages and facets of the process". (Peel, 1971)

"Designing curricula in a way that reflects the basic structure of a field of knowledge requires the most fundamental understanding of that field". (Bruner, 1960)

Specification of educational objectives presupposes such knowledge (Bloom, 1956). Ausubel et al (1978) reflect these premises but, like Iansek and Balla (1979) from consideration of the role of the teacher:

"It seems self evident that the teacher should constitute an important variable in the learning process. From a cognitive standpoint it should certainly make a difference, in the first place, how comprehensive and cogent the teacher's grasp of his or her subject matter is". (p. 498)

In the pedagogy of the diagnostic thinking process, that subject matter concerns the cognitive operations of both neophyte

and experienced clinician. Therefore, it is considered necessary for the present research to take samples of subjects at more than one station along the developmental continuum and so students, house officers and registrars are studied. De Groot (1965) and Marton (1975) describe the reasons for adopting a developmental and comparative approach:

"... the more 'experience' a person has collected in any field, the more difficult it becomes for him to understand the behaviour of have-nots. Thus, every teacher knows the following frequent brand of overestimating his students: opining that from the given problem situation his students can 'immediately' derive (see) some property or means that he himself finds quite obvious - whereas in reality, in order to 'see' it, much perceptive and abstractive experience is required. The teacher has had this experience for so long that he is no longer aware of it. An experienced problem solver in any field is particularly apt to forget about his primary and fundamental problem transformations even before he starts his own consciously operational thinking". (de Groot, 1965)

"A necessary pre-requisite of ease of learning on the part of the learner ... is that the teacher has a clear conception of what lack of understanding looks like and as regards the absence in the student of these pre-requisites on which it may depend ... The idea is simple enough: in order to help the students understand, we must first understand their way of thinking about the topics with which we are concerned". (Marton, 1975)

It is shown in subsequent chapters that, despite much work, our knowledge of the diagnostic thinking process and its development remains at a rather general level (see section 3.1 in particular) and that by this the development of pedagogical approaches is hampered (see section 4.2.3).

A second major reason for the present research is that the study of the development of the diagnostic thinking process in medical education and clinical practice has implications not only for the medical student's training, (Schroder et al, 1967) but also for the efficiency and effectiveness of his subsequent practice. It is not suggested that the experienced clinician's diagnostic thinking processes are without blemish.

Hence, achievement of a better, more complete and accurate description of the process than is currently available might allow advantage to be taken of Abercrombie's (1960) hypothesis. She substantiated by her research that "we may learn to make better judgments if we can become aware of some of the factors that influence their formation". This is given greater significance by Barrows and Mitchell (1975) who point out that "central to the effective delivery of health care by the physician is the complex skill of clinical problem solving". Knowledge of their own thinking processes should enable clinicians to modify them according to the demands of each individual clinical situation. Such knowledge may also stimulate a constructive questioning of established applications and a more rigorous self evaluation.

A final, and no less important, reason for the present research is suggested by Norman et al (undated, unpublished report). This is that the study of clinical problem solving is of significance to psychology and psychologists as an example of general problem solving behaviour, and the development of an appropriate theoretical framework for teaching and learning is of significance to the body of educational theory and its applications.

1.2 Aims and Outline of the Research

The research aims of the present study are twofold:

Firstly, to add increased clarity, specificity and breadth to current descriptions of the diagnostic thinking process in undergraduate medical education and clinical practice.

Secondly, to propose possible pedagogical strategies in relation to the efficient and effective development of appropriate diagnostic thinking processes.

To achieve these aims complementary research methods were developed. The first comprises two structured, multiple choice questionnaires in endocrinology and neurology, respectively. The questionnaires were designed to analyse the relative contributions of different types of knowledge towards predicting

subjects' diagnostic ability and to compare these in two separate specialities of internal medicine. Each questionnaire is made up of four sections dealing with different aspects of the diagnostic process:

- A. Mastery of factual knowledge.
- B. Interpretation of symptoms and signs.
- C. Selecting and testing diagnostic possibilities.
- D. Formulating a diagnosis.

The questionnaires were each administered to two groups of subjects for comparison of performance at the end of undergraduate medical education and after some years of clinical practice. The subjects were final year clinical medical students and medical registrars.

The second research method is a form of introspective account gathering by stimulated recall of clinical interviews using immediate videotape playback. The subjects for this part of the study were final year clinical medical students, pre-registration house officers and medical registrars. This part of the study allows comparison of the three groups' diagnostic thinking processes per se but does not lend itself to the rigorous quantification possible with the multiple choice questionnaires. The two research methods are, however, complementary, each lending interpretative value to the results of the other. From them an overall psychological perspective on diagnostic thinking is achieved (see Chapter Twelve).

The research design allows analysis of the developmental aspects of the diagnostic thinking process and of the relative effects of medical education and clinical practice. Taken in conjunction with a review of current teaching approaches and a discussion of related theoretical issues in teaching and learning, the pedagogical implications of the research findings are drawn.

1.3 Areas Excluded from the Research

Some closely related areas and issues are not within the

bounds of the present research. These may be confused with its proper subject matter which is the diagnostic thinking process of subjects of varying clinical experience and their implications for medical education. The present study is therefore confined to the realms of cognitive psychology and education. However, because of the potential for confusion, we may usefully mention some of the major excluded areas.

Firstly, the study does not address itself to either the interviewing or interpersonal skills of the students and clinicians who form its samples of subjects, regardless of whether or not such skills might enhance data gathering. We address ourselves only to the information with which each subject worked, whether that information had been either actively elicited or offered by the patient without prompt, and to the cognitive manipulation of it. The subject's handling of the patient while gathering and processing the information is not a variable for this study.

Secondly, the study was not designed to measure subjects' competence in terms of factual recall in the manner of, for example, Byrne and Freeman's (1971) study. A test of factual knowledge is included, but this is strictly circumscribed, and is not generalisable to the subjects' level of professional factual knowledge as a whole. Clinical problem solving obviously has an element of retrieval of information from long term memory, but the study concentrates on the subjects' cognitive processing of the information so retrieved (as well as other information) rather than measuring the size of the store or efficiency of the retrieval mechanism as such.

Thirdly, we do not discuss whether or not clinical enquiry and problem solving may be seen as a scientific method (see, for example, Medawar, 1969; Dudley, 1970; Forstrom, 1977). Barrows and Mitchell (1975) suggest that the skill of the physician is not an art but "an exacting discipline, every bit a scientific method in its own right". It would seem a matter of philosophical debate, however, whether or not such cognitive

processes may be characterised as either art or science. It is not a debate in which we wish to join.

Fourthly, the study does not attempt to define diagnosis, a term about the meaning of which there is some lack of clarity or consensus (see, for example, Ruff and Mechanick, 1975). Although our subjects were asked to make a diagnosis where they could, they were left to define for themselves what constituted that diagnosis. In other places, we have taken diagnosis to mean that which appeared under such a heading in the patient's medical record. A precise definition of the term would have hampered those aspects of the study which required subjects to define where and when they had reached a concluding point and is of no interpretative value in our consideration of a process of thinking.

Finally, the study is confined to hospital medicine. We have no indication of the generalisability of our findings to general practice medicine, which is different in a number of aspects (Thwaites, 1954; Royal College of General Practitioners, 1972).

1.4 Content and Organisation of the Thesis

The content and organisation of the thesis may be presented diagrammatically as follows:

<u>CHAPTER</u>	<u>CONTENT</u>	<u>GROUP</u>
2	Mathematical and logical models	
3	Cognitive studies	Literature review
4	Pedagogy, theory and practice	
5	Hypotheses, methods, subjects, validity	Overall research design
6	Validity, reliability, data analysis	
7	Results	Questionnaire study
8	Discussion	

<u>CHAPTER</u>	<u>CONTENT</u>	<u>GROUP</u>
9	Method, validity, reliability, data analysis	
10	Results	Account gathering study
11	Discussion	
12	Unified psychological perspective	
13	Pedagogical implications	Synopsis
14	Limitations and indications	

1.5 Summary

Major reasons for the current research are described and summarised. Aims of the research are defined and an outline of methods and subjects given. Areas and issues excluded from the study are listed and the content and organisation of the thesis explained.

CHAPTER TWO

Review of the Literature 1. Mathematical and Logical Models

The diagnostic thinking process has been characterised in a number of ways which may be divided into two types; firstly, are those models of the process which are based on statistical, algebraic or logical approaches, and, secondly, are those models and theories which have a basis in cognitive theory. In each case, some work has been based in research, and some in little more than speculation. Although our own interest does not lie primarily in mathematical or logical models of the process, these must nonetheless be reviewed since they achieve some considerable currency in the literature and appear to attract some credit as representations of the actual thinking process. This chapter, then, will give a brief description of such models and will evaluate their usefulness in illuminating the diagnostic thinking process. Chapter Three will give a more detailed account of cognitive approaches, these constituting our major interest.

2.1 Statistical Models

The development of statistical models of the diagnostic process has occurred in parallel, and with considerable interaction and mutual influence, in the fields of medicine and clinical psychology. The following discussion will, therefore, include work in the latter field where this is relevant to the former.

Goldberg (1968) traces the development of statistical modelling of the diagnostic process from early research on accuracy of clinical judgments in clinical psychology, and validity studies to determine in what way clinical judgments are accurate. Such investigations progressed to the process of clinical inference and the representation and simulation of clinicians' cognitive processes. The statistical models which have been, and are being, developed predictably fall into the two broad categories of linear and non-linear. In turn, the studies can again be sub-divided into two groups as indicated by Hoffman et al (1968); those which focus upon outcome, reliability and accuracy,

and those which focus upon the judgment process itself especially upon the manner in which cues are weighted and combined by the clinician.

2.1.1 Linear Models

Typical of this approach have been models based on linear regression, analysis of variance and conditional probability or Bayes' Theorem. Typical of the linear regression approach are the studies described by Hoffman (1960) and Hammond and Summers (1965), although these are in clinical psychology. In such models, the dependent (or criterion) variable is the clinician's diagnostic judgment, and the independent (predictor) variables are the values of the cues (test scores, symptoms, signs, etc.). The result of such an analysis is a set of regression weights, one for each predictor. In such studies, clinicians are asked to estimate some criterion on the basis of given values of several predictors. Hoffman (1960) used a multiple regression procedure to compute the relationship of each predictor variable to the decisions of each judge across a group of test protocols. The results, stated in terms of 'relative weights' indicate the relative importance of each predictor in contributing to the decisions of each clinician and to the criterion classification. Complex terms may be inserted into such linear models, in an attempt to account for non-linear relationships in the data. Hoffman's (1960) model included in the equation cross-products between pairs of variables.

Hammond and Summers (1965) cite more than a dozen studies of clinical judgment in which the accuracy of prediction derived from linear regression analysis was great enough, for them, to suggest that clinicians are primarily linear in their mode of combining cues. Although they do not eliminate the possible existence of meaningful non-linear cue use, they considered that it had not yet been exhibited in the experimental situation. Three years later, Wiggins and Hoffman (1968) compared three models as representations of the cognitive processes of

29 clinicians, and, although for some clinicians a non-linear model was slightly better, for 23 of the 29 a linear regression model was equal or superior to a quadratic model.

Closely related to linear regression models are those which are based on analysis of variance. Such methods became known as 'policy capturing'. Using analysis of variance, Hoffman et al (1968) captured the policies of radiologists in diagnosing the benign or malignant nature of gastric ulcers. When clinicians' judgments are analysed in terms of analysis of variance, a significant main effect for Cue 1 would imply that clinicians' responses varied systematically with that cue. A significant interaction between Cues 1 and 2 would imply that clinicians were responding to particular patterns of these cues. In general, then, such a model represents clinicians' judgments as a weighted, additive combination of the factors, both main effects and interactions. Significance tests can provide an indication of the clinician's use of a particular cue individually or in combination with one or more other cues (Hoffman et al, 1968).

Hoffman et al's (1968) results indicated statistically significant differences among their nine radiologist judges, both in diagnoses made, and in their use of individual cues and cue patterns. However, groups of judges appeared to agree with one another. Four types of judge were identified by principal components analysis and analysis of variance indicated the nature of the differences between these groups in terms of the cues used. However, in view of the small number of subjects in this study (nine) one may question the validity of the statistical analysis and results.

Analysis of variance has proved less useful than linear regression as a method of developing statistical models of the diagnostic process for two reasons. Firstly, its use is limited to those situations for which experiments can be designed so that all cue patterns are orthogonally arranged within a factorial design. This is not possible for more than a small

number of cues. Secondly, by orthogonalising the set of cues which are correlated in nature, one might alter the judges' conceptualisation of the environment, thereby causing him to deviate from his normal judgment model. This second problem - a lack of congruence with reality - is a major drawback and jeopardising variable for almost all studies concerned with the development of statistical, algebraic or algorithmic models.

The final major approach to the development of linear models has been based on conditional probability theory or Bayes' Theorem. According to this, in relation to medicine, accurate decision-making depends on the prior probabilities of the possible diagnoses and the observed signs and symptoms, and on their probability of joint occurrence. The strength of an association is indicated with probabilities also. Bayes' Theorem allows mathematical revision of opinion about possible diagnoses in the light of new information.

The earliest, and probably most formative paper on the use of conditional probability models in computer-assisted diagnosis is that of Ledley and Lusted (1959). This paper laid a broad foundation in symbolic logic, conditional probability theory and statistical decision theory on which most of the subsequent work in this area has been based. Taylor (1971) gives a comprehensive review of conditional probability models of diagnosis. Taylor's (1970a) own opinion, reflected by Card (1970), is that in clinical practice, the doctor collects data sequentially, guided at each stage by a mental estimate of the probability of diseases under consideration, and his Bayesian model of this process showed a success rate of 93 per cent. However, the model had to select between only three possible diseases, which introduces a certain element of unreality into the process; firstly, by having the possibilities designated and provided, and, secondly, by limiting them to only three. This element of unreality is present in all such studies, and was recognised by Taylor (1970b). However, his solution of developing a system with more tests (88), nine diagnoses and eight treatment possibilities does not necessarily reduce the unreality of being

'given' possibilities from which to select in the first instance. The technique still demands a closed set of symptoms and possible diagnoses (Lindberg, 1968), and it is in this that the unreality is seated, not in the magnitude of that closure.

Nonetheless, Bayesian conditional probability models of the diagnostic process have been widely used. Taylor (1971) quotes studies in which Bayes' Theorem has been applied to problems in haematology, gastro-enterology, cardiology, primary bone tumours, psychiatric diagnosis and Cushing's syndrome, in addition to his own work on thyroid disease. Knill-Jones et al (1973) used a Bayesian model in diagnosis of jaundice, and Knill-Jones (1977) reports its use in calculating the prognosis for severe head injury patients. Lusted and Stahl (1964), however, point out that Bayesian models have greatest diagnostic accuracy when used in areas such as thyroid disease and congenital heart disease in which diagnostic data is drawn mainly from laboratory tests rather than the clinical history and physical examination which is difficult to present in quantitative form. Lusted and Stahl (1964) conclude their paper with the suggestion that perhaps the Bayesian model does not entirely account for the clinician's thinking process. They speculate that the diagnostic thinking process is more nearly one of pattern recognition than numerical adjustment, although, after pattern recognition, the clinician may bring in his own subjective probabilities, developed from past experience and related to 'pattern weighting'. There is no evidence to support such a view.

De Dombal et al (1971) take this further. On the basis of their finding no statistically significant correlation between levels of certainty reached by probabilistic analysis and certainty levels recorded by clinicians, they conclude that clinicians do not 'think Bayes' or that pure probability theory plays a relatively small part in the diagnostic thinking process. Gorry (1968) also disagrees with those who consider Bayes' Theorem to model the clinician's thinking processes.

He points out that we cannot, as yet, capture the processes used by physicians in a sufficiently precise manner to permit use in a computer programme, and that such a simulation of human thinking processes would fail to exploit the special capabilities of the computer. Despite this, two years later Gorry (1970) suggests that Bayes' Theorem is a reasonable model of the physician constantly updating his 'current view' of the patient's problem, with the qualification that "a basic model of the diagnostic process may be valid without accounting for the detailed processes employed by individual doctors".

A number of other workers have considered the possibility of clinicians thinking in terms of probability estimates (Sarbin et al, 1960; Meador, 1969) and, indeed, the Royal College of General Practitioners (1972) considers that all diagnoses are themselves statements of probability; a view reflected by Ledley and Lusted (1959). However, the opposing viewpoint, that clinicians do not think in terms of exact probabilities, is held by other workers (Albert, 1974) and substantiated in other fields (Anderson and Shanteau, 1970).

These, then, are the major linear models of the diagnostic process, and reference to the papers cited will evidence an associated consistent and considerable degree of predictive accuracy. Despite this, linear models have been the target of criticism. Hoffman's (1968) discussion mentions some of the points of controversy. Firstly, adoption of a linear model would imply that individuals do not alter their mode of 'weighting' the dimensions of information, regardless of their pattern or configuration. Secondly, clinicians "report in fairly emphatic terms that judgment involves a sequential consideration of many dimensions (symptoms, signs or cues), and that the interpretation of a given dimension is conditional upon the values of other dimensions. This subjective assessment is supported by Goldberg (1968), and mentioned by Meehl (1954, 1960). Elstein et al (1978) found the very diagnostic accuracy of the linear models which they

applied to their data reason enough to reject the linearity principle:

"Thus to the extent that a linear model resulted in increased diagnostic accuracy without changes in the data base, its adequacy as an account of human performance may be questioned". (p.104)

In clinical psychology, Nystedt and Magnusson (1975) quote five studies between 1968 and 1970 alone which indicate that judges use cues in a configurative way, typical of which is that of Wiggins and Hoffman (1968) which demonstrated 'configurality' as a consistent judgmental characteristic of 16 of 29 judges. In view of such factors, some workers have addressed themselves to the development of non-linear models, considering that the interpretation of symptoms and signs is conditional upon the presence, absence or nature of other symptoms and signs.

2.1.2 Non-linear Models

As Hoffman et al (1968) indicate, there have been very few studies of the degree to which judges operate in a configural manner as opposed to using some weighted combination of individual cues and those there are may be described in the framework of Brunswik's Lens Model Theory or probabilistic functionalism.

For a detailed discussion of Lens Model Theory, see Postman and Tolman (1959); but briefly, the model has, unlike others so far mentioned, a large cognitive component. It is used to describe the process of clinical judgment by describing the judgment situation as an interaction between two systems: an ecological system (being the outer environment) and a cognitive one (being the judge, or clinician, and judgment or diagnostic process). (Figure 2.1 gives a simplified representation of the Lens Model, applied in a clinical situation).

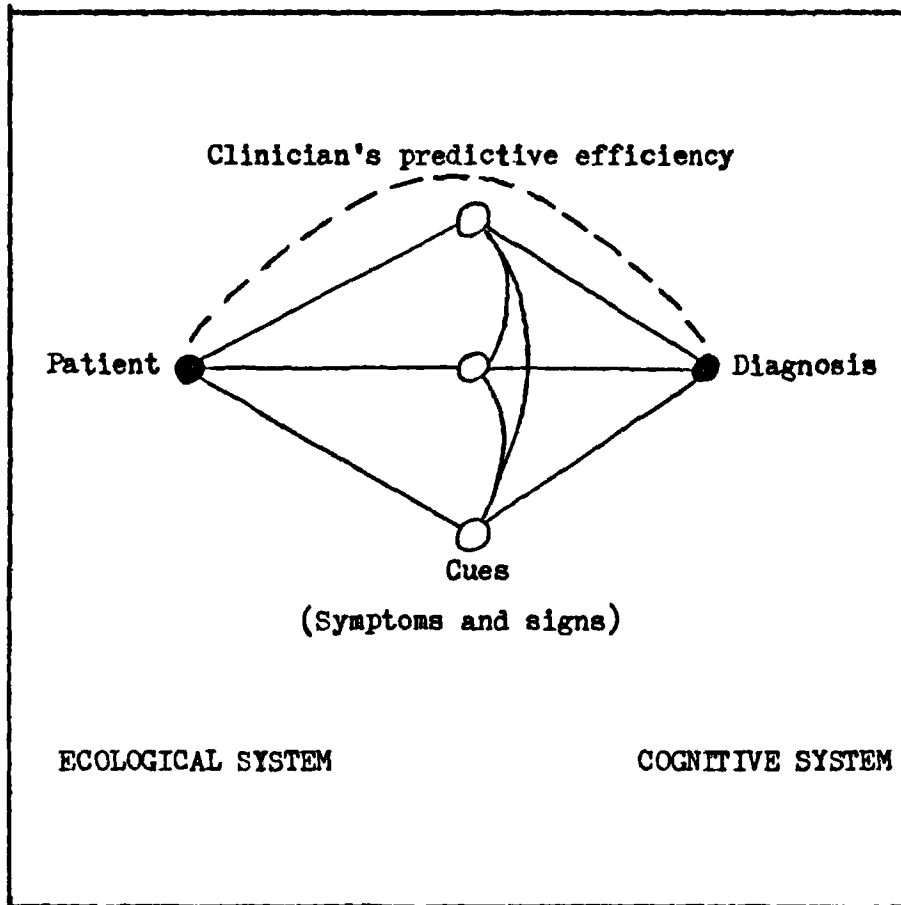


Figure 2.1: Simplified Representation of the Lens Model

The cognitive system may be seen as a mediatory link between the input - providing ecological system and the instrumental response system. The unbroken lines in Figure 2.1 represent correlation coefficients or probabilities. It is seen as necessary to consider both the systems of ecology and cognition since qualities within the ecological system restrict the clinician's possibilities of making perfect predictions, while properties within the cognitive system limit his maximal predictive efficiency.

The Lens Model will not be discussed further since, although it has been widely applied in the field of clinical psychology

(Sarbin et al, 1960; Goldberg, 1970; Nystedt, 1972; Brehmer, 1973) no study has yet been completed in the field of medicine. It therefore remains for a non-linear model to be developed.

2.2 Other Models

Prominent among other models has been the use of decision trees, decision analysis and decision theory. Schwarz et al (1973) discuss the nature and use of decision trees and associated probabilities and values. They give an example of a decision tree describing possible actions by the physician, and their potential consequences, in a patient thought to have either essential hypertension or functionally significant renal artery stenosis. The tree consists of nodes and branches. At decision nodes the physician must choose one from a set of actions and proceed to travel down the consequent branch to the next node. Although Schwarz et al (1973) suggest that "most physicians will find the diagrammatic representation quite in keeping with their thinking about medical problems", there is no experimental or research evidence of such a thinking process. Some decision trees and flow-charts, such as those of Essex (1976), do not purport to represent a thinking process, but are merely an aid to diagnosis, as are others reported by Taylor (1971). Kleinmuntz (1968) has used logical decision trees to study diagnostic behaviour of clinical neurologists.

Studies using the theoretical framework of decision theory and decision analysis have been completed by Garland (1959) with radiologists, comparing the effects of lax and strict decision attitudes and Aitchison (1970) in relation to treatment allocation. Aitchison and Kay (1973) interpret the clinician's diagnostic behaviour in terms of decision theory and the reduction of uncertainty. However, to what extent each of these represents a formal model of the diagnostic thinking process is not always made abundantly clear by the authors.

The only remaining major interpretive framework - again, it is difficult to judge the extent to which authors consider these to be models of the diagnostic process - is that based on set theory. Feinstein's (1967) book has been formative in this area, and is clear that a model is being proposed:

"Like the character in Molière's "Le Bourgeois Gentilhomme", who was astonished to learn that he spoke in prose, clinicians may be startled to discover that they think in mathematical sets. The thinking occurs during every act of diagnosis, prognostic estimation, therapeutic decision, and correlation of clinical and laboratory data. As exercises in deductive and inductive reasoning, these acts can be described in mathematical terms".
(p.156)

Such an attempt to describe the clinician's thinking processes had been made previously (Feinstein, 1963), yet the discussion of the application of set theory shows that a clinical taxonomy, or an objective organisation and classification of clinical data, is being achieved, and not a description of a thinking process. Set theory and Venn diagrams were attractive in this endeavour, because of their ability to identify and construct overlapping collections of items. As Bashook (1976) suggests, "the logic Feinstein presents is reasonable and attractive except that it represents what we would like to see in physician thinking and not what our present knowledge suggests as reality". The Royal College of General Practitioners (1972) also has used set theory as an interpretive framework, but without suggesting that clinicians actually think in this way, while Dudley (1971) suggests that clinicians "run through Boolean lattices of a diagnostic continuum".

These, then, are the major statistical and logical models of the diagnostic process. As has been indicated, authors do not always make it clear whether or not their formulations are intended as formal models of a thinking process, and have not usually been evaluated as such experimentally. Despite this,

the very construction of models, particularly of statistical models, has given rise to considerable controversy. A discussion of this controversy may indicate the degree to which credibility may be attached to the models described as descriptions of a thinking process.

2.3 Statistical versus Clinical Prediction

The essence of this debate turns on whether, in a given prediction situation, the procedures of the clinician or the statistical actuarial formulae are more accurate. Although this question per se is not of relevance to our study, the elaboration of arguments associated with its answer is of interest in evaluating the models already discussed.

When clinician and actuary have been compared, the latter has almost invariably been found to be as good as or better than the former in accuracy of clinical prediction or diagnosis in clinical psychology. Meehl's (1954) work set the scene. Having surveyed the field, he concluded that, given the same quantitative or scalable information about the patient, statistical methods of combining it were consistently as good as or better than clinical methods in diagnosing the correct disease. Sawyer (1966) later summarised results from 45 comparative studies and gave strong support to Meehl's 1954 conclusion.

Similar comparisons of clinician and statistical formula have been made in medicine, and usually show similar results (de Dombal et al, 1972a; Lusted and Stahl, 1964). Taylor et al (1971) have not only compared outcomes, but also the diagnostic processes of man and computer. Where the choice is to be made between three possible diagnoses, the probability estimates as they are revised during the process, may be plotted within a triangle and the 'pathways' to diagnosis compared. (See Figure 2.2)

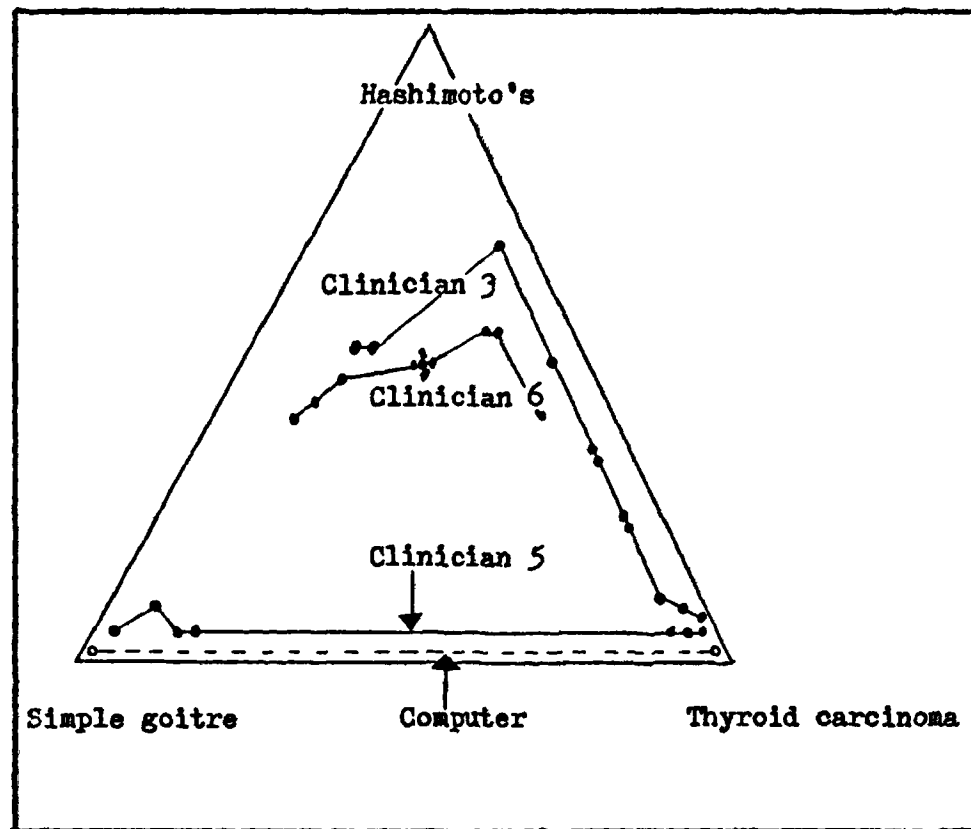


Figure 2.2 Illustration of Clinicians' and Computer's
Pathways to Diagnosis, Plotted within a Triangle
(after Taylor et al 1971)

The pathway selected by the computer is normally shown to be the most direct.

On the basis of such findings as these in clinical psychology, Sarbin et al (1960) have maintained that, fundamentally, clinical prediction is always actuarial, and since this is the case whether the clinician knows it or not, it is to his advantage to make these actuarial predictions explicitly so. However, it is the arguments against this viewpoint which provide the basis of an evaluation of all mathematical and logical models of the diagnostic thinking process.

Shulman and Elstein (1975) review a basic distinction between 'process-tracing' approaches, which attempt to describe the intellectual processes used by subjects forming judgments,

making decisions or solving problems, and the 'black box' investigations which attempt to model the processing of the judge mathematically through studies of input-output relations. The models discussed in this chapter are primarily of the 'black box' variety. Even Taylor et al's (1971) 'pathways' fall into this category, giving no indication of how probability estimates are arrived at, or whether, were the subjects not forced to behave like the computer, they would naturally do so. Elsewhere, Elstein et al (1972) point out that statistical models replicate the judgments of the clinician without necessarily reproducing his mental steps.

Holt (1958, 1970) argues that a comparison of clinician and statistical model is wrongly set, since the issue of how data is gathered is ignored, and only combination of data in isolation from other processes considered. Sawyer (1966) supports this view. One can add to this that even that combination of data is not always as it may be in reality. Taylor et al's (1971) work gives a fine example of clinicians competing with a computer, on the computer's terms. The allocation of numerical, rather than subjective, probabilities, even if the concept of a numerical probability is understood, is not conspicuous in everyday problem solving thinking, and evidence has yet to be provided that it is present in everyday diagnostic thinking processes. Even Sarbin et al (1960) admit to differences between clinical and actuarial processes. Dudley (1970) writes of "a sense of odds ... derived from other methods of inference", but not a sense of numerical probability.

It would seem reasonable to conclude that comparisons between clinician and statistical formulae are spurious. The two are not compared on equal terms. The true comparison is not between statistical and clinical methods, but between a statistical model and a clinician attempting to behave as a statistical model, having been given an array of data in an unfamiliar form and an unfamiliar context and being allowed to use clinical skills hardly at all. It is not to be wondered at, then, that he should compare unfavourably with a mathe-

mathematical formula. It is interesting to note that the primacy of the processes of the computer rather than the clinician in developing a model has been questioned by some workers in the field and Pauker, Gorry, Kassirer and Schwarz (1976), some of whose work has already been cited, have tried to tackle the problem by taking the clinician as primary and attempting to make a computer simulation of his processes, rather than vice versa.

A final evaluation of mathematical and logical models may be made by reference to Hoffman (1960), who considers the problem of whether models are paramorphic or isomorphic representations of the clinician. That is, whether they are symbolic representations or exact replicas of the clinician's thinking processes. If a model using, for example, linear regression analysis, derives an accuracy of prediction equivalent to that of a computer, we cannot logically conclude that clinicians are primarily linear in their mode of combining. We would contend that each of the models discussed is a paramorphic representation of the diagnostic process, not an isomorphic one. In order to understand the diagnostic thinking process, we must consider work which takes a study of the clinician as its major reference point.

2.4 Summary

Mathematical and logical models of the diagnostic process are described. These have been based on linear and non-linear statistical methods, decision theory and set theory. The relative effectiveness of the clinician and the models is discussed. It is concluded that, regardless of relative predictive validity, mathematical and logical models are only symbolic representations, and not exact replicas of the clinician's diagnostic thinking processes.

CHAPTER THREE

Review of the Literature 2. Cognitive Studies

In making a psychological enquiry into the nature of the diagnostic thinking process, two alternative approaches are available. Either the process as a whole may be traced and characterised, probably in fairly broad terms; or variables of possible relevance to that process may be identified, studied separately and the results of such studies be used in an additive manner. Examples of the former approach are seen in the application of such conceptual and explanatory frameworks as 'pattern recognition', 'information processing' and 'hypothesis generation and testing'. Examples of the latter approach are seen in studies of the effects of varying amounts of clinical experience, redundant information and variability in content. Of course, another choice of approach is available in that the process as a whole may be broadly defined and the effects of isolated variables on that process studied either consequently or simultaneously. This final approach has been the selection of the two most significant groups of workers in the field - Barrows and his colleagues at the Project for Learning Resources Design at McMaster Medical School, Canada, and Elstein, Shulman and their colleagues at Michigan State University.

This chapter will begin with a review of the work of the groups at Michigan and McMaster which may be characterised under the broad heading of "Hypothesis Generation and Testing". This section will include a report of their work on both the characterisation and description of the process as a whole as well as indicating which variables they have isolated for special study and the results of those studies. It is to these studies that the present work, to be reported, is most closely related. The aim of the first section is specifically to describe the findings and theories of these two most significant groups, only scant reference, therefore, will be made to either complementary or contradictory findings from other workers. The full review of other work will be made in subsequent sections and, where appropriate, related to the findings of the McMaster and Michigan groups. Likewise, where variables isolated and studied by these two groups have also been studied by other

workers, they will be reviewed under separate headings in later sections.

3.1 Hypothesis Generation and Testing

Although the findings and viewpoints of the Michigan and McMaster groups are very similar, we shall commence with a separate description of each, referring in each case to recent publications in which each group has drawn together and reviewed its own findings and described its own current thinking.

3.1.1 The Michigan Medical Inquiry Project Studies

These are collected and summarized in Elstein et al (1978). The book reports a series of separate, but related, studies by different members of the group. Their most realistic method of enquiry comprised three high-fidelity simulations in which actors were trained to simulate conditions in haematology, gastroenterology and neurology. The subject interviewed and examined the simulated patient in the normal manner, during which time he was encouraged to verbalise his thoughts as much as possible. This included episodic reviews during the interview. The clinical interview was recorded on videotape. The videotape was then replayed to him. He was given a stop/start switch with which to control the playback and was encouraged to use the videotape as a vehicle to stimulate his memory and relate what he was thinking throughout the session being reviewed.

This process resulted in three kinds of data. Firstly, material from concurrent thinking aloud; secondly, episodic reviews during the clinical interview; and thirdly, material from videotape - stimulated recall. Where discrepancies were found they were weighted highest to lowest in the order given. Although verbal report protocols have been widely used and have achieved a respectable place as a method in psychological research (Peel, 1971; Newell and Simon, 1972), the technique of 'thinking aloud', which Elstein et al

(1978) imply to be their most important source of data, has been criticised (de Groot, 1965) for its disturbing influence. It is reported that thinking aloud during a problem-solving exercise has the effect of markedly slowing down the thinking itself and obliging the subject to think more explicitly, 'at a higher level of conscious organisation, than he would otherwise do. In other words, thinking aloud may itself alter the thinking process of which it is a report. Even though Flaherty (1975) found no differences in problem-solving scores while thinking aloud, Elstein et al's (1978) results of the high-fidelity simulations must be conservatively judged in the light of this qualification.

The second research method employed by Elstein et al (1978) involved the use of paper and pencil simulations, called 'patient management problems', in which the subject was presented with a brief verbal description of the patient's problems and then recorded his steps towards making a diagnosis. Each step involved the selection of a test or examination or further information, in response to which a decision about the next step may be taken. Results of these tests, again, must be interpreted conservatively in the light of the unrealistic nature of the "patient", the data, the means of eliciting information and the necessary structuring of the subject's strategy and tactics of problem-solving.

In a third study, fixed-order problems were presented to subjects in which the sequence of data presentation was predetermined. Four such cases were presented, which varied along two dimensions, diagnostic specificity and cue consistency. Two cases were designed to converge on a single diagnosis, whereas in the other two several cues could reasonably be applied to more than one diagnosis. Further, in only two problems were all cues consistent with one diagnosis. This method, obviously, represents a low-fidelity simulation, but was designed to study the interaction between the structure of the diagnostic problem itself and the physician's own problem-solving strategy.

The subjects for the three studies were some or all of 24 experienced, peer-nominated physicians, who were subdivided into two groups: 'criterial' physicians who were judged by their peers as good clinicians, and 'non-criterial' physicians who had not been so judged, but no differences were found between these subjectively derived groups on any measure. Parametric statistical analyses were applied to some data, although the number of subjects may appear to be inappropriately small. This must be borne in mind when evaluating Elstein et al's (1978) results.

Finally, before reporting results, a brief description and discussion of the scoring system and analysis of data from the high-fidelity simulations is necessary in order to evaluate the validity of results. We may take Elstein et al's (1978) own summary of the system, which took more than a year to develop:

"The fundamental units of protocol analysis are information search units, cues and hypotheses. Information search units tabulate the data-gathering behaviour of the physician; cues are the data obtained; and hypotheses are the formulations of possible solutions to the problem. A matrix of cues and hypotheses was constructed to represent the appropriate cue weightings for each hypothesis, as determined by an expert. The analysis of each protocol initially examined the following variables: total number of information search units, point at which the first hypothesis was generated, number of hypotheses active one-quarter and one-half of the way through each work-up, total number of hypotheses generated, number of hypotheses retained at conclusion of work-up, number and per cent of cues acquired, number and per cent of critical findings obtained, efficiency, accuracy of interpretations, modal interpretive error, accuracy of formulation, and accuracy of outcome". (pp 62-63)

It is the case that data obtained by verbal report protocols cannot be taken as complete. De Groot (1965) quotes the opinion that "a protocol is relatively reliable only for what it positively contains but not for that which it omits". In addition, a verbal report protocol, particularly if elicited by Elstein et al's (1978) second and

third methods of episodic reviews and stimulated recall, cannot be taken as a simultaneous commentary on thought processes; thus identification of time of cognitive occurrences may be hazardous. These two deficiencies in verbal report data render Elstein et al's (1978) variables less valid than their quantitative nature may imply. Despite their assertion that "a conservative line of interpretation of the stimulated recall segments was consistently adopted", it would appear, on the contrary, that some over-interpretation of data has occurred. This also is to be considered when evaluating the results of their studies.

Elstein et al (1978) summarise their findings in a four-stage general model of medical inquiry that calls attention to cue acquisition, hypothesis generation, cue interpretation, and hypothesis evaluation. This model has also been discussed elsewhere (Elstein et al, 1977; Shulman and Elstein, 1975). Findings showed thoroughness of cue acquisition and accuracy of cue interpretation to be statistically independent. The clinical interview is interpreted as consisting partly in a guided search for findings implied by each hypothesis, although Elstein et al (1978) were unable to distinguish reliably between hypothesis - testing questions, and questions asked routinely. In the process of cue interpretation, data are evaluated in terms of their fit to the anticipated findings. It is suggested that this evaluation is ordinarily conducted by designating each cue as positive, non-contributory or negative with respect to a particular hypothesis. This latter finding, however, is based on the comparison of differential weighting systems with the result that, "the 'pro', 'con' and 'does not help' formulation does at least as good a job of describing information - processing behaviour as does a more complex differential weighting system" (p.100). This would suggest a paramorphic model, rather than a description of a cognitive process (~~Elstein and Sprafka, 1974~~).

The findings further suggest that diagnostic accuracy is related to both thoroughness of cue acquisition and accuracy of cue interpretation, despite these two variables being uncorrelated. The majority of diagnostic decisions could be accounted for by one of two rules: either select the hypothesis with the maximum number of positive cues; or, select the hypothesis with the maximum difference of positive cues minus negative cues. However, for one problem, rules proved superior to physicians in diagnostic accuracy, which may well, again, suggest a form of paramorphism. Elstein et al (1978) have defined other judgmental rules as: "If feature A is absent, it cannot be diagnosis X", while some diagnostic strategies are implied by the very definition of a disease: "X is diagnosed if and only if all of the following features are present". However, no indication is given of the frequency or circumstances of application of such rules, the possible range of rules, the consciousness or not of their applications, or each clinician's degree of awareness of them. It may be more appropriate to describe such circumstances not as 'rules' but as a logical description of cognitive strategies displayed. The term 'rules' seems to have too many implications and defining parameters (definiteness, method of application) which cannot be accounted for.

A further finding is that hypotheses are generated early in the clinical interview, Elstein et al (1978) suggest by ten per cent of the way through or, usually, within the first five minutes of the interview. In earlier work (Elstein et al, 1972) these early hypotheses had been identified as specific rather than general. However, this formulation is now rejected. The nature of the hypotheses are now seen as dependent upon the structure of the medical problem. They may therefore be specific or general at any stage of the interview. Kleinmuntz's (1968) finding of a 'general to specific' development may, therefore, not be generalisable but specific to the neurological cases in question or to the research method used. In addition to hypotheses being generated early, new ones may be generated later in the

interview or older ones reformulated. As described by Sprafka and Elstein (1974), "the hypotheses define the framework of possible solutions to the diagnostic problem and thus constitute a problem space (Newell and Simon, 1972)", transforming the ill-defined problem into a well defined one.

The concept of a problem space is taken from information processing theory and amounts to describing the subject's internal representation of the task environment. According to the theory, the structure of the task environment determines the possible structures of the problem space, while the structure of the problem space determines the possible programmes that can be used for problem solving.

The clinician's choice of hypothesis is obviously not arbitrary or random. Elstein et al (1978) consider that they are typically generated by associations from clusters of a few cues. Each hypothesis accounts for some, but usually not all, of the initial data base. They find that the number of hypotheses considered at any one time is limited, averaging between four and seven and rarely exceeding five - although it must be noted that the limitations of verbal report protocols may also limit the number of hypotheses reported. Elstein et al (1972) give the figure as 4 ± 1 , and their derivation as having four components - attending to initially available cues; identifying problematic elements from among these; associating from problematic elements to long-term memory and back, generating hypotheses and suggestions for further enquiry; and, informally rank-ordering hypotheses according to subjective estimates. The concept of 'problematic elements' has since been replaced by cue clusters.

An important finding of Elstein et al's (1978) work has been that clinical competence may be case related, both physicians and students varying considerably in their diagnostic effectiveness, according to the nature of the problem in hand. In

particular, intra-individual consistency across the patient management problems was found to be low, although this finding is strongly qualified by noting that the task environment is itself structured by the problem solver into a psychological problem space. Despite this, they suggest that "among people who are knowledgeable about a class of problems, they are structured in predictable ways, adapted to the task". Content-specificity of the problem solving process is seriously questioned by Berner and Bligh (1974) who found more similarity between students in performance across cases on the same skill than within a given case on different skills, the skills being: definition of an initial problem list, correct diagnostic procedures, incorrect procedures chosen, and final diagnoses. It is possible, however, that a group of experienced physicians would show different results from these students.

We may put forward an alternative, and reconciling, interpretation of these apparently conflicting findings. It appears that the measures of the problem solving process made and compared across cases are measures of a process per se to different degrees, some being measures which are directly related to content, in particular to amount of information, and diagnostic significance of cues. For any given amount of information, or given number of cues, there is a finite number of possible diagnoses; this number will vary from cue cluster to cue cluster. Likewise, cues differ in their diagnostic significance; presentation of 'a pain in the chest radiating down the left arm' is likely to give rise to fewer diagnostic hypotheses and a narrower initial range of questions than 'a pain in the chest'. Elstein et al's (1978) variables are largely content-dependent in this manner, it therefore follows that content-specificity should be displayed. This is quite in accord with the information processing proposition that the structure of the task environment determines the possible structures of the problem space. The problem solving process of hypothesis generation

and testing, however, is common to all problems and subjects studied. This may more appropriately be called a process than the other quantitative parameters. This viewpoint is echoed by Berner et al (1977) who also consider that such studies do not allow process and content aspects to be separated.

Although Elstein et al's (1978) studies are not primarily concerned to compare expert and neophyte, indication is given that information and experience are basic to competence in clinical problem solving.

Finally, the study indicates certain sources of error in the clinical problem solving process. There are dangers inherent in allowing hypotheses to influence data collection and interpretation. These dangers include possibilities of premature closure, inappropriate selective information gathering, and biased interpretation of information. Other errors identified include mistakes in combining evidence, misinterpretation of single cues, and faulty hypothesis generation. In a non-medical context, Shulman and Elstein (1972) quote Wason's (1968) work, suggesting that a hypothesis may be a 'strong expectation' and may become a self fulfilling prophecy by the failure to gather or process information that has a negative weight for a favoured hypothesis. Other errors noted include over-interpretation and under-interpretation of cues, excessive data collection where patients are implicitly assumed to have multiple problems, and uninterpreted cues.

Elstein et al's (1978) findings, then, have been quite precise and specific. Some methodological problems which may jeopardise validity of those results have already been mentioned. In addition to these, we may ask whether the studies have addressed themselves to the best questions, or whether they have been far-reaching enough. Emphasis on quantitative methods has largely limited the studies to

certain quantifiable aspects - counts and weightings of cues and hypotheses, divisions by time intervals, etc. - which are essential but not necessarily sufficient. We know, therefore, how much information is collected, and how it relates to hypotheses, but we have no indication of the clinician's cognitive manipulation of cues, how these are cognitively structured as clusters, and what process of structuring generates hypotheses. How is it possible for more than one hypothesis to be generated? What is the cognitive mechanism which admits of simultaneous (or alternating) interpretations of cues? How do new hypotheses arise? It is stated that the first hypothesis is generated by ten per cent or the first five minutes of the interview. What cognitive processes are occurring during the nine per cent or four minutes preceding the emergence of the hypothesis? How are relationships between cues allocated and reallocated? According to what tactic or strategy is the clinical interview conducted prior to generation of the first hypothesis? These questions are not yet answered, although the McMaster group have approached some of them.

3.1.2 The McMaster Studies

The findings of the McMaster group are summarised in a recent, unpublished report (Barrows et al, 1978). It will be seen that this work closely reflects the findings of the Michigan group, and refers to a study of 37 physicians, made up of 19 general internists (medical registrars) and 18 family physicians (general practitioners), in 62 standardised patient encounters, using simulated patients as in Michigan.

In a previous pilot study (Barrows and Bennett, 1972) it was found that physicians developed multiple hypotheses early in the clinical encounter (cf. Michigan studies). These hypotheses (guesses, hunches, impressions or possible diagnoses) were seen as a guide for what was essentially a problem based inquiry. It must be noted that the McMaster work is within the frame of reference of the problem-oriented approach.

ach to clinical practice and teaching and the development of problem-based learning. The 1972 study also showed that many physicians in residency training had ineffective problem solving skills. The authors suggested that efficient problem based performance is not developed during undergraduate or postgraduate training, but appears in clinical practice 'when the physician encounters heavy patient responsibilities with insufficient time to handle his responsibility in the traditional way he had been taught'.

For the 1978 study, each of the 37 physicians was asked to examine one of four randomly allocated simulated patients in the normal manner, but being advised to take about 30 minutes. A year later two-thirds of the original group (13 general internists and 12 family physicians) participated in a second phase. There was, therefore, a total of 62 physician-patient encounters, from 37 subjects on four simulated patients. The research method differed slightly from the Michigan design, omitting any thinking aloud.

Each physician-patient encounter was monitored by two observers. A clinician noted down ideas about what the subject might be thinking on the basis of his questions and physical examination. These observations were compared with the subject's subsequent statements about his own thinking. The second observer recorded inter-personal relationship behaviours. The encounter was videotaped with the image of a digital timer superimposed. At the end of the encounter the subject dictated notes about the patient and then reviewed the videotape, being asked standardised, non-directive questions, particularly about thoughts which prompted his questions and physical examination. When the subject expressed hypotheses, he was asked how they were ranked, or whether they changed during the encounter. The videotape was stopped "whenever discussion seemed appropriate". Significant points during this discussion were recorded by an observer, as were the physician's actions on history and examination seen on the videotape. The digital timer was

used to co-ordinate all observations. A typed transcript was then prepared containing the physician's questions and actions, the patient's responses, and structured observations concerning his and the monitor's observations. The physician was then asked to rate the effect of the information he obtained in supporting or denying the hypotheses he expressed on a +2 to -2 scale. They also rated whether each question and item of physical examination was 'routine' or 'non-routine'.

The information thus obtained was coded into 21 variables for computer analysis (coding procedure is described in Norman et al, 1977). Results showed that 57 per cent of time was spent in history taking, 12 per cent in physical examination and 21 per cent in combined activity. On average, 61 per cent of questions were non-routine, that is directed at a definition of the problem. About half the questions asked were specifically to test hypotheses. Subjects reported that routine questions were used for activities such as scanning for unrelated and unsuspected problems, building rapport, or to gain thinking time. On average, physicians elicited 68 per cent of the available significant information.

The extent to which a finding was based on multiple hypotheses was calculated by expressing the average number of hypotheses against which a finding was weighted positively or negatively. On average, a finding was relevant to about three hypotheses. Barrows et al (1978) conclude that physicians were using parallel rather than sequential processing. This is not a permissible conclusion, however, since any one finding could be processed repeatedly for hypotheses arranged sequentially, and the relationship to multiple hypotheses merely be a function of the inter-relatedness of the hypotheses due to their common derivation from the same information. The average weight of findings suggested that the physicians tended to elicit data that confirmed rather than denied hypotheses.

Multiple hypotheses, then, emerge as a central feature in physicians' thinking (cf. Michigan studies). In contrast

with the Michigan findings, however, the first hypothesis in these studies was advanced, on average, within 28 seconds (range across physicians one to 262 seconds) of the appearance of the main complaint. However, the timing of that event for each encounter is not given. The correct hypothesis emerged, on average, four minutes into the encounter (range across physicians one second to 32 minutes). We see, then, a large degree of variability across physicians. In both cases, however, variability across cases was less (cf. previous discussion of content-specificity). On average, physicians advanced 5.5 hypotheses (range across physicians three to nine). The range shows a slight discrepancy with the Michigan results. Seventy-nine per cent of hypotheses were generated within the first quarter (by time) of the encounter.

In their discussion, Barrows et al (1978) suggest that generation of the first hypothesis, which appears in a matter of seconds, results from a process of pattern recognition in which the initial cues are matched with similar instances from the physician's prior knowledge and experience. This, however, does not account for the formation of first hypotheses which appear much later, as some did. Neufeld et al (1976) defined possible sources of hypotheses as anatomical structure relating to the symptom, recall of a specific previous patient and classified text book information. But this is speculative. Early hypotheses may be general or specific (cf. Michigan studies). The generation of multiple hypotheses is seen as preventing the clinician from prematurely closing on an acceptable but incorrect diagnostic conclusion. This does not preclude, however, the possibility of some hypotheses being 'favoured', while others are treated more lightly (see discussion of Michigan studies).

In their broad view of the process of clinical reasoning and enquiry, the McMaster work supports that of the Michigan group. It is a dynamic process of multiple hypothesis generation which guides enquiry for hypothesis testing. The

McMaster studies add further quantitative data, but their validity is subject to the qualifications applied to the Michigan studies. The McMaster group have also added a further dimension to the study by comparing subjects with varying degrees of clinical experience.

Norman et al (1977) found no differences in time of onset of first hypothesis (60 seconds) or number of hypotheses (six to seven) between beginning medical students and a sample of physicians. First-year students elicited, on average, 47 per cent of available information, compared with 69 per cent for final-year students, and 64 per cent for physicians, although accuracy of diagnosis was lower (33 per cent for first-years, 67 per cent for final-years, 100 per cent for physicians). Neufeld et al (1976) found that students, like physicians, used the method of hypothesis generation and testing. As with physicians, about half the questions asked were specifically to test hypotheses. Rimoldi (1964), however, found physicians' questions more directly case related than those of students, and the number of questions asked decreased with experience. Leaper et al (1973) report similar findings as does Kleinmuntz (1968). Barrows and Bennett (1972) suggest that students and house staff tend to invest their hypotheses with considerable precision and specificity, "whereas good clinicians tend to keep their hypotheses broad and vague - allowing them to be shaped and categorised by the data from the patient" (p.276). However, Barrows (1976) states that their studies have shown that "a physician's experience with previous problems affects the richness, the appropriateness and the usefulness of the hypotheses that he develops" (p.24). At the same time, Neufeld et al (1976) suggest that a model of effective problem solving is characterised by, amongst other things, the use of fairly specific hypotheses. The question of the contribution of experience in determining the nature and usefulness of hypotheses, then, seems to remain unanswered.

In conclusion, the McMaster studies throw more light on the question of the clinical reasoning process, but many questions remain unanswered. It seems that, having identified a process which may be labelled "hypothesis generation and testing", this very label has limited research to such matters of qualitative description as How many? When? and Related to what information?. The question of What else besides? has not been put or answered. The questions put at the end of the previous section, therefore, remain.

To have gathered research evidence of a process of hypothesis generation and testing is an important step, but perhaps not as remarkable as the literature may suggest. Ordinary, adult thinking processes are of this hypothetico-deductive nature (Inhelder and Piaget, 1958; Peel, 1971; Bruner, 1973) and it has been maintained that children also function in this manner (Shapson, 1977). It would seem reasonable that no special broad mechanisms of thinking are necessary to the clinician, but rather that ordinary adult thinking processes are applied to a special, and specially problematical, content. It is the precise mechanism of that application and the nature of the cognitive operation of that process which requires further illumination.

3.2 Pattern Recognition

Barrows et al (1978) and Scadding, (1967) suggest that generation of the first hypothesis occurs as the result of a process of pattern recognition. Such a process has been invoked by a number of writers, either in partial or in total explanation of the diagnostic process. For example, Hamilton (1966) suggests that making a diagnosis is a comparison of one disease 'profile' with a standard 'profile' which is an average of all information available to physicians in the past. Therefore, not all previous patients, nor the present one, had that exact profile. Pattern matching, therefore, must be 'as close as possible' and, of necessity, will have an element of probability about it. Gorry (1970) holds a

similar viewpoint. Hubbard in Jacquez (1964) holds a similar opinion that much of diagnosis is pattern recognition. However, a major question is how a minimum number of adequate points that are acceptable in order to identify a pattern is to be identified. Sprafka and Elstein (1974) have represented the process of medical diagnosis as a list-matching activity in which the physician compares observed findings to the contents of lists retrieved from memory. Such a process would also involve questions of probability and sufficiency. These latter problems have been approached by workers considering pattern recognition in relation to computer applications in diagnosis (Kulikowski, 1970; Lissack and Fu, 1976).

It is to be noted that no evidence of a pattern-recognition process in medical diagnosis is available, it is merely put forward as an explanatory concept, with greater or lesser degrees of certainty. For example, the Royal College of General Practitioners (1972) suggests, according to the essentially Hippocratic method (Scadding, 1967), that the doctor "records a pattern which can be compared with patterns in a textbook or remembered from previous clinical experience. From these, the one most closely resembling the present case is chosen to be the diagnosis", whereas the Learning Resources Design Project (Programme for Educational Development, 1975) at McMaster suggest that diagnosing by pattern recognition is fraught with the danger of forcing the patient to fit the pattern.

Evaluating the theories of a process of pattern recognition is made hazardous by the lack of definition of the term by those who use it, and the concentration of psychological research on visual and auditory pattern recognition. However, we may say that if, and in whatever form, it takes place, it is not a process of template matching, if only because of the uniqueness of each patient. In essence, the clinician is presented with what may be called a partial pattern and

must recognise this as a possible instance of a pattern, or disease, already known. However, these partial patterns do not really present themselves even as simple partial patterns, free of distractors. For example (Lindberg, 1968), the onset of diabetes mellitus classically consists of excessive thirst, excessive appetite, copious urine and weight loss. When a patient presents his major problems as loss of libido, grey hair, weight loss and insomnia, somehow the clinician must construct the appropriate partial pattern as well as hypothesise or infer the entire one. He must initially operate on that cue cluster in some way before pattern recognition of any type is possible. In what sense, then, might the process of reaching a diagnosis have elements of pattern recognition in it?

It would seem probably to be only in a very peripheral manner, in that some of the same principles may hold true. We may cite the principles that, in terms of predictability, although the whole is greater than its parts, this is not to imply that the whole is unpredictable from its parts; and that the subject's perception may depend upon his expectations of the total set of possible patterns (Corcoran, 1971). However, although such principles may hold true in both pattern recognition theory and diagnostic decision-making, they probably do so for entirely different reasons. In terms of visual or any sensory pattern recognition each element of the pattern may have no meaning in its own right independent of all other items in the pattern whereas this does not apply to any one symptom, sign or item of the patient's history. Each may have a meaning, or range of meanings, in its own right. In addition, a group of such symptoms, signs or items of history has no meaning in its own right, unlike, for example, three lines, two of which join at one end while the third is laid across them. Meaning, that is to say a diagnosis, can only be attributed to an array of clinical information by means of some other referent - such as pathology of an organ or a system, or the action of a virus or bacteria. Pattern recognition, as such, therefore, does

not appear to provide a very useful explanation of diagnostic thinking but may obscure more subtle and complex cognitive processes. For example, we may cite Peel's (1971) work on bringing outside information into consideration when forming judgments, or Bruner's (1957) discussion of going beyond the information given. For Peel (1971) the characteristic of the highest level of judgmental responses to given information is that of conceiving of possible hypothetical contingencies which have to be evoked from the thinker's own experience and insights. Any theory of an element of pattern recognition in the diagnostic thinking process must, at best, be secondary to considerations such as these.

3.3 Confirming, Refuting and Irrelevant Information

Kleinmuntz (1968) used a form of the 'Twenty Questions' parlour game with neurologists and from it defined 'diagnostic trees'. After having proceeded about half way down a particular tree, subjects were asked to recall or recite all information that they had accumulated prior to that point. It was found that they did not remember those data that did not substantiate a particular differential diagnostic hypothesis. In other words, they seemed selectively to forget "irrelevant" data; irrelevant, that is, to their hypotheses, although possibly not irrelevant to the true diagnosis. This penchant for confirming evidence was also noted by Smedslund (1963) who showed that when nurses were presented with a series of cases in which a particular symptom was associated with a particular diagnosis as often as it was not, the subjects concluded that the symptoms and diagnosis were positively correlated. As discussed above, when a hypothesis serves as a 'strong expectation', negatively weighted information may not be processed. Sprafka and Elstein (1974) found that a diagnostic predictive model based on negatively weighted cues performed less well than models based wholly or partially on positive weights.

This tendency to over emphasise positive information is

well documented in the psychological literature. Wason (1968) produced compelling evidence, using numbers games, that "even intelligent individuals adhere to their own hypotheses with remarkable tenacity when they can produce confirming evidence for them The subjects appeared to display rigid or fixated patterns of behaviour because they failed to overcome the set created by their confirming evidence" (p.172). Peel (1971) also discusses this phenomenon on the basis of his research findings. He states that once a particular hypothesis has been devised or accepted, the thinker has difficulty in shifting to another. This difficulty is twofold; firstly, the thinker is more likely to try to hold on to his first view, even if it has to be modified, in the face of new observations, and, secondly, new explanations may be obstructed "by the directive and selective influence of the accepted theory upon the subsequent collection of observations and results. People, even learned ones, may tend to look for what supports their theory and consciously or unconsciously neglect evidence which contradicts it". (p.74). Such data Peel calls 'hypothetico-observations'. A number of possible reasons for the tendency to confirm rather than refute hypotheses are forwarded by Peel (1971): firstly, the organic quality of thesis development as opposed to a clear-cut process of logical elimination; secondly, confirmation of one hypothesis may, in some circumstances, imply negation of the other; thirdly, the thinker does not have to go beyond positive thinking; fourthly, mental inertia may predispose a person to neglect to clear away alternatives; finally, we tend to judge things in our environment by their positive presence.

Reitman (1964) considers constraint proliferation in the solution of ill-defined problems and suggests that "as additional constraints are added, the problem solver's investment in some increasingly particular and delimited area of the hypothetical solution space mounts. For many

problem solvers, the sheer fact of this cost sunk in a particular approach will suffice to make exploration of any other unlikely" (p.299).

Finally, the tendency to verification is further evidenced by Evans and Wason (1976), while Donaldson (1959) gave some explanatory evidence that subjects found it more difficult to work with negative rather than positive information and that the two, although logically equivalent, may not be psychologically equivalent. However, her subjects were 14 year olds and Inhelder and Piaget (1958) identify this as a stage in the development of formal operational thinking. In a correlational task concerning the relationship between eye and hair colour, at Substage III-A their subjects could not relate confirming and non-confirming cases and the set of all possible cases. Subjects did not appreciate that $p.q$ and $\bar{p}.\bar{q}$ are both favourable to a general statement about the relationship of p and q . Instead, only $p.q$ is considered. By Substage III-B, however, this is resolved, the subject seeing certain cases as being the inverse of others, and understanding the reciprocity between confirming and non-confirming cases. However, presuming that the subjects of the clinical studies cited have achieved Substage III-B, it would appear that psychologic is sometimes more dominant than formal logic in problem solving thinking.

The question of information load and relevance has been approached by some workers. Scadding (1967) points out that irrelevant features, unless identified as such, may obscure diagnosis not only by suggesting false conclusions, but also by falsely limiting the range of possibilities considered. The British Medical Journal (1977) cites studies of damage to doctors' performance due to the constant flow of large quantities of often irrelevant information in a short space of time. Studies are cited which show that doctors may overlook radiological and bacteriological evidence of active tuberculosis, fail to react to adverse drug effects and

miss a host of other laboratory 'clues' to diagnosis, while in some circumstances clinical diagnostic accuracy seems to be related to the amount of information elicited. Nystedt (1974) and Nystedt and Magnusson (1972) found similar adverse effects on judgment with increasing information load in clinical psychology. Psychological research supports this finding, while Jones et al (1978) add that subjects do not normally shift to different strategies in order to deal with the increasing load, and Streufert's (1973) work suggests that complex decision making varies with relevance, while simple decision making varies with information load. Peel (1971) also finds that irrelevance of information is often a stumbling block to sustained reasoning.

It would seem reasonable to suggest that efficient operation under varying conditions of information load and relevance is related to the subject's knowledge base and to his set of hypotheses, the latter being somewhat dependent upon the former. It has been suggested (Learning Resources Design Project, 1975) that a well designed 'net' of hypotheses should encompass all items of information, and thus assist memorisation by 'chunking' the information. Such a formulation allows no room for any piece of information to be identified as irrelevant. We are therefore left with a mechanism which can cope with high information load where the clinician's knowledge and experience base is sufficient and appropriate, but cannot account for the possible need to identify information as irrelevant thereby averting the possibility of generating irrelevant hypotheses and overloading and decreasing the efficiency of the problem solving process. The question of how irrelevant information is to be identified during the problem solving process therefore remains.

3.4 Other Factors

In clinical psychology, studies have been made of the relationships between redundant and new information, clinician's confidence and time available for assessments

(Huff and Friedman, 1967), and clinicians' confidence levels in relation to amount of information (Oskamp, 1965), but no systematic studies have been reported in medicine, although Elstein (1976) mentions the phenomenon of collection of redundant information in order to bolster confidence in a clinical inference. Simborg et al (1976) quote characteristics of the patient, the practitioner and the clinic setting as factors shown to influence clinicians' decisions, and add the recording of information to that list themselves. But these are not studies of thinking processes as such.

Two related factors which have received a little attention are the subject's representation of information in both long-term and short-term memory, and the role of these stores in the diagnostic process. Kleinmuntz (1968), on the basis of his 'Twenty Questions' games in neurological diagnosis postulates, firstly, that the neurologist's search strategies involve the use of both short-term and long-term memory storage, the former undergoing moment-to-moment modifications and revisions, evidenced by his subjects' ability to remember only those data relevant to particular diagnostic hypotheses. Secondly, he postulates that the neurologist has a visual representation of the central nervous system in his long-term store and that much of his information processing involves shuttling backwards and forwards between symptoms, test findings and neuroanatomical locus. Kleinmuntz's evidence for this latter postulate is rather sketchy, being based solely on the subjects' discussion of anatomical loci. In addition, since 'localisation of the lesion' is a major feature of neurological diagnosis, we cannot speculate on the generalisability of the postulate. Indeed, Elstein et al (1978) conclude that the mode of mental representation involved in the generation of problem formulations is, for all physicians, predominantly verbal, and that, for most individuals, mental imagery only occasionally occurs, but is used as an adjunct to the primary verbal mode. However, to what extent this conclusion is a function of the verbal mode of account gathering, and the nature of the case

involved we cannot speculate. The question of the subjects' representation of information in both long-term and short-term memory remains an important one, with considerable implications for medical education and for the ability of the individual clinician to control and evaluate his own hypothesis generation and testing behaviour, particularly if hypotheses are generated, as Elstein et al (1978) suggest, by an associative process that links cues to content stored in memory. The Michigan group elicited conflicting results about whether these cues are single or in clusters, and suggest that differences in research method may have caused this conflict, as yet unresolved.

Hypotheses also may serve a purpose in overcoming the limited storage capacity of short-term memory, in that they may serve as rubrics in short-term memory under which cues are stored (Sprafka and Elstein, 1974). This point is also emphasised by the Learning Resources Design Project (1975). Human problem solvers work under severe limitations of memory and processing capacity, which are compensated for by such strategies (Reitman, 1964).

Finally, the effect of personality on diagnostic thinking processes was considered by Elstein et al (1978). A small battery of tests was administered to their subjects, including measures of logical reasoning, dogmatism, flexibility and cognitive complexity. Quoting their own assessment of results: "The association between personality variables and clinical problem solving measures was inconsistent at best". There were also no statistically significant correlations between measures of logical problem solving and clinical problem solving. As mentioned above, formal logic and the logic of the psyche are not necessarily one and the same. However, one study has reported statistically significant correlations between personality scores and ratings of clinical performance by peers and supervisors (Kegel-Flom, 1975), although the interaction effects between these two measures may be considerable. Klein et al's (1969) non-medical work on anxiety and learning to formulate hypotheses suggests

that there are no anxiety-related differences in the number and quality of hypotheses generated, but their more anxious subjects set higher standards, did not report those evaluated as too poor, and so communicated fewer of a higher quality than less anxious subjects. This might suggest that less anxious subjects would benefit from training in hypothesis evaluation.

3.5 Summary

The two main series of research studies of the diagnostic thinking process (the Michigan and McMaster studies) have revealed a process common to students and clinicians which may be defined as "hypothesis generation and testing". The specific research methods of the Michigan and McMaster groups are described. Both employ simulated patients and videotape - stimulated recall, while the Michigan group add the techniques of thinking aloud and episodic review, as well as patient management problems. Scoring systems and quantitative parameters are described. Both research studies are criticised for spurious quantitation and over-interpretation of inexact data. ,

The major conclusions of the Michigan group may be summarised as follows:

- 1) A four-stage general model of medical inquiry is postulated. This consists of cue acquisition, hypothesis generation, cue interpretation and hypothesis evaluation.
- 2) Cue acquisition and accuracy of cue interpretation are statistically independent.
- 3) Clinical data are evaluated in terms of their fit to anticipated findings. Each cue is designated as positive, non-contributory or negative with respect to a particular hypothesis.
- 4) Diagnostic accuracy is related to thoroughness of cue acquisition and accuracy of cue interpretation.
- 5) Most diagnostic decisions could be accounted for by one

of two rules; either, select the hypothesis with the maximum number of positive cues; or, select the hypothesis with the maximum difference of positive cues minus negative cues.

- 6) Hypotheses are generated early in the clinical interview. They may be specific or general and further hypotheses may be generated as the interview progresses.
- 7) Hypotheses define the framework of possible solutions to the diagnostic problem and are generated by associations from clusters of a few cues.
- 8) The average number of hypotheses held at any one time is between four and seven.
- 9) Clinical competence may be case related. This finding is questioned and an alternative interpretation put forward.
- 10) Information and experience are basic to competence in clinical problem solving.
- 11) Possible errors in the clinical problem solving process are identified as premature closure, inappropriate selective information gathering, biased interpretation of data, mistakes in combining evidence, misinterpretation of single cues, faulty hypothesis generation, over-interpretation, under-interpretation and uninterpretation of cues, and excessive data collection.

Many of these conclusions are questioned. It is considered that the data do not support such precise findings. In certain instances, also, the statistical analysis is found wanting (see section 8.10 below). In addition, many relevant questions are neither put nor answered. These are identified as:

How are cues cognitively manipulated?

What process of structuring generates hypotheses?

How can multiple hypotheses be generated?

What cognitive mechanism enables multiple interpretations of cues?

How do new hypotheses arise?

What cognitive processes occur before generation of the first hypothesis?

How are relationships between cues allocated and re-allocated?

According to what tactic or strategy is the clinical interview conducted prior to generation of the first hypothesis?

The major conclusions of the McMaster studies may be summarised as follows:

- 1) Physicians develop multiple hypotheses early in the clinical encounter.
- 2) Efficient clinical problem solving skills develop with clinical practice rather than medical education.
- 3) Physicians spent 57 per cent of the clinical interview in history taking, 12 per cent in physical examination and 21 per cent in combined activity.
- 4) On average, 61 per cent of physicians' questions were non-routine. About 50 per cent in both physicians and students were hypothesis testing questions.
- 5) Routine questions are used for scanning, building rapport and to gain thinking time.
- 6) On average, physicians gain 68 per cent of available significant information.
- 7) Each clinical finding was relevant to about three hypotheses.
- 8) The correct diagnostic hypothesis emerged, on average, four minutes into the encounter (range across physicians one second to 32 minutes). Variability across cases was less than variability across physicians.
- 9) On average, physicians advance 5.5 hypotheses (range three to nine) 79 per cent of which are generated within the first quarter of the encounter.

- 10) Generation of the first hypothesis arises from a process of pattern recognition. This theory is considered to be inadequate.
- 11) Early hypotheses may be general or specific.
- 12) Multiple hypotheses prevent premature closure.
- 13) Some hypotheses are 'favoured'.
- 14) Students and physicians do not differ in time of first hypothesis or number of hypotheses.
- 15) First year students elicited 47 per cent of available information; final year students 69 per cent; and physicians 64 per cent.
- 16) Accuracy of diagnosis was 33 per cent for first year students; 67 per cent for final year students; and 100 per cent for physicians.
- 17) Students and physicians display the process of hypothesis generation and testing.

It is concluded that the McMaster studies are subject to the same problems of spurious quantitation as the Michigan studies. The McMaster conclusions complement those of the Michigan group, but the unanswered questions, identified above, remain largely unanswered due to the limiting effects of the theory of "hypothesis generation and testing". This formulation is evaluated in the light of the work of Piaget, Bruner and Peel, and is not found unexpected.

Theories of a diagnostic process of pattern recognitions are considered and rejected as over simplified in their present form since they do not fully take into account the partial nature of the clinical information offered and elicited. Studies of the effects of confirming, refuting and irrelevant information are discussed and related to broader psychological theory. The question of how information is designated as irrelevant remains unanswered.

Other factors which may influence clinicians' decisions are

discussed, including the roles of short-term and long-term memory, the internal representation of information, and personality. Very little research has been conducted in these areas.

3.6 Implications for the Present Research

The literature review has revealed a number of points and omissions which are important for the present research design and associated hypotheses. These points and omissions are as follows:

- 1) The application of methods of data collection and interpretation of the Michigan and McMaster groups have been severely criticised. Although no criticism of the method of stimulated recall qua research method is offered, it is considered that the Michigan additions of thinking aloud and episodic review seriously jeopardise the validity of the data. It is also considered that the counting or recording of exact timing and numbers of hypotheses on the basis of stimulated recall data is unmerited since we may not rely on either completeness or congruent timing of reports. In addition, some findings seem to be interpreted beyond the tolerance of the associated data. For example, the McMaster studies showed that a clinical finding, on average, was relevant to about three hypotheses, and the conclusion is drawn that physicians use parallel rather than sequential processing. It is argued (above) that such a conclusion is not permissible because, firstly, any one finding may be processed repeatedly and sequentially in relation to different hypotheses and, secondly, the relationship to multiple hypotheses is less likely to be a deliberate economy of cognition than a function of the necessary inter-relatedness of different hypotheses derived from the same data. Our present research design, then, must ensure that similar errors of research method and data interpretation do not occur.
- 2) Neither the Michigan nor McMaster groups offer a full

discussion of the theoretical background or rationale associated with the method of stimulated recall, which is necessary to enable proper evaluation of the reasonable scope of the method. The present study will offer such a discussion and evaluation (see Chapter Nine).

- 3) Areas of omission as identified above will be addressed by the present study (see research hypotheses, section 5.1 below).
- 4) The literature review has identified or is suggestive of certain definable components of the diagnostic process, the subject matter and the person. These are knowledge, cue interpretation, hypothesis testing, arriving at a diagnosis, generalisability or specificity of thinking processes and amount or nature of clinical experience. Conclusions related to these variables, as yet, are not without some degree of doubt. The present study, therefore, will begin with a study of these defined variables, based on multiple choice questions and case histories. This will provide a strong, appropriately quantified background against which further qualitative findings gathered from accounts by stimulated recall may be more accurately interpreted and some of the present dark areas illuminated (see research hypotheses, section 5.1.. below).
- 5) Finally, although some relevant work has been referred to above, no consistent comparison of student and clinician, or of the effects of medical education and clinical practice, has been made. The present study, therefore, will, at all stages, compare student and clinician and provide an indication of the developmental aspect of diagnostic skill.

CHAPTER FOUR

Review of the Literature 3. Teaching the Diagnostic Thinking Process

Since the mid-1950s, attempts have been made to devise teaching strategies appropriate to the development of the diagnostic thinking process. Although these have occurred against a background of massive innovation and change in medical education, it is not our intention to review or discuss this broader context, but merely to describe the main responses when, as Elstein et al (1978) express it:

"With increasing frequency medical educators were told that their objective was to produce problem solvers, inquirers, individuals skilled in gathering and interpreting information for the purpose of rendering judgments, making decisions and taking action". (p.2)

Of course, it cannot be said that before this time no attempt was ever made to teach students how to make a diagnosis, but it can be said that the attempts were neither as explicit nor as systematic as those to be described here.

However, these approaches have largely been purely empirical in their development and implementation or have been based on the hypothesis generation and testing model and the concomitant problem centred approach to learning. So far, there has been lacking any evaluation of these teaching strategies against a background discussion of models or theories of teaching and learning. Before describing and discussing the various teaching methods devised, then, a discussion of relevant aspects of theories of teaching and learning will be presented.

4.1 Theoretical Aspects of Teaching and Learning

Any discussion of teaching and learning can be pursued at a variety of levels of specificity. For example, criteria for the selection of teaching method may be defined so that a rational choice may be made between lecture, group discussion, practical classes, self-instruction, educational games or simulations, problem solving exercises and so on. Or the discussion may be in terms of teaching media: Should we use

film, slide, audiotape, videotape, models, handouts or some other aid? We may consider teaching tactics such as provision of feedback, opportunity for active learning, repetition and reinforcement of learning, assessment strategies and remedial teaching. Or we may be interested in teaching styles: authoritarian, heuristic, socratic or counselling. The present discussion, however, addresses none of these since decisions or recommendations about them can only follow on from specific knowledge of the characteristics of the learners, the teachers, the material to be learned and the context of teaching and learning. Instead, questions of a broader, perhaps more fundamental and generalisable, nature will be considered.

We take as our starting point Peel's (1971) conclusion that "the level of judgment may be quite susceptible to cultural and educational influences" and consider some of the major, general conditions which may promote mature and reasoned judgment in the medical student or, at least, provide a firm foundation for its future development. We take it also that the diagnostic thinking process may be defined sufficiently to make it amenable to a planned teaching strategy. However, it is our contention that such definition is, as yet, lacking in appropriate specificity. So, other criteria must also be used to assess and evaluate the teaching methods devised. In particular, we shall consider three fundamental concepts which are seen to be relevant given the degree of specificity already achieved in defining the diagnostic thinking process, and the range of reasonable generalisations to be made in applying teaching and learning theory to particular situations. These concepts are: structure in learning; transfer of learning; and the relationship between problem solving and learning.

4.1.1 Structure in Learning

The concept of structure in learning may be interpreted in a number of related ways. It is difficult to consider any one of these without implicit reference to or acceptance of the

others. This difficulty arises from the necessary set of inter-relationships between the objective, internal structure of the subject matter to be learned, that structure as internalised, either similarly or differently, by teacher and learner (which may be referred to as cognitive structure) and the structure of presentation of the subject matter by the teacher and its reception by the learner. These structures may complement or conflict with important results. For example, McKeachie (1963) and Gage (1964) cite research showing that students who structure the field in the same way as the teacher, even though they do not agree with the teacher's opinion, complete courses more successfully than do other students. This finding is given support by Broadbent (1975).

However, McKeachie expresses uncertainty about whether or not a teacher could communicate his structure to students who did not already have it. Perhaps, however, this would be an unreasonable aim if, as Bruner (1960) describes it:

"Grasping the structure of a subject is understanding it in a way that permits many other things to be related to it meaningfully. To learn structure, in short, is to learn how things are related".

Since these relationships and manners of relating may vary depending upon the prior knowledge and structures of the individual, then perceived subject structures or cognitive structures may also vary, but be equally efficient in later use. Bruner emphasises this aspect of learning also:

"Students, perforce, have a limited exposure to the materials they are to learn. How can this exposure be made to count in their thinking for the rest of their lives?"

This must surely be the central question when devising strategies for teaching the diagnostic thinking process also.

Bruner's emphasis on thinking is particularly apposite, for the diagnostic thinking process has a structure, whether the same or different from clinician to clinician or student to

student or student to clinician, and the content of that process and its referents (previous knowledge) also have a cognitive structure of their own. Definition of these structural aspects is complicated by the ill defined nature of the diagnostic problem and the dynamic, gradually unfolding nature of the diagnostic problem solving process. Thus structural aspects of the content of that process, if the model of hypothesis generation and testing is conditionally accepted, may change many times during its progress, referring to a series of hypotheses which may be stable in their individual structural characteristics. In other words, the process may be one of cognitive structuring and restructuring of information. We may also be certain that the diagnostic thinking process itself is a structured one if we take H.C.A. Dale's reasonable definition cited by Bartlett (1958). According to this definition, " a system is unstructured if it is such that search for the assigned objective can just as well begin at any stage in the system as any other". Although this is substantively true, the direct implication that "there is indeed no possible objectively preferential order" for search cannot be true once the first piece of clinical information has been elicited and the problem solver's response required. Thus the relationship between structure of content and the process of restructuring is a complex one. Teaching for that relationship affords even greater complexities, for, as Lunzer (1968) and Broadbent (1975) point out, an experimenter (or teacher) may define a situation as an objective stimulus complex, but the subject's (or learner's) schemata (prior knowledge) determine what will function as a stimulus for him, that is, what he will attend to, what he will retrieve from memory, and so on. This brings us back again to the question and role of cognitive structure in relation to curriculum design.

Shavelson and Stanton (1975) define cognitive structure as "a hypothetical construct referring to the organisation (relationship) of concepts in long-term memory". They consider

that education attains the goal of communication of subject matter structure to the learner effectively and efficiently when "there is a close correspondence between the subject matter and the representation of the subject matter structure in the cognitive structures of students". Again, with reference to teaching the diagnostic thinking process, this correspondence is a complex one. The complexity arises from the dual nature of the subject matter itself for, on the one hand, it has a definable content (factual information, explanation, skills, interpretation, classification and so on) while, on the other hand, based on this content are processes (making a diagnosis and managing the patient) the development of which must be a major aim of both undergraduate and post-graduate medical education. In a sense, the major subject matter of medical education may well be taken as the relationship between the content and process of clinical medicine. These separate aspects, as well as their inter-relationship, must find proper representation in the student's and teacher's own cognitive structures. In the event, we may find these considerations to be complex only in philosophical terms. In pedagogical practice, these complexities may resolve into relatively simply teaching strategies.

For Bruner (1966), a theory of instruction, and presumably its practice too, must, amongst other things:

"specify the ways in which a body of knowledge should be structured so that it can be most readily grasped by the learner. Optional structure refers to the set of propositions from which a larger body of knowledge can be generated, and it is characteristic that the formulation of such structure depends upon the state of advance in a particular field of knowledge The goodness of a structure depends upon its power for simplifying information, for generating new propositions and for increasing the manipulability of a body of knowledge". (p.41)

These final two criteria could equally well apply to the diagnostic thinking process itself. Perhaps theories of teaching as cognitive restructuring (Gage, 1964) seem particularly relevant here, especially in the light of Bruner's (1964)

additional contention that "knowing is a process, not a product". The learning and practice of medicine provide an example par excellence of this statement, and, indeed, such a concept is not foreign to its British halls (Womersley et al, 1974). In essence we are echoing Svensson (1977) who contends that "the structural properties of an individual's conception of various phenomena are the most important aspects of his knowledge, because they also constitute fundamental characteristics which determine how he acquires and uses knowledge". However, we add that the conception here considered requires in turn an analysis and understanding of a series of structures and structural relationships which may vary according to the context of the individual concerned.

4.1.2 Transfer of Learning

The concept of transfer of learning concerns whether or not, by learning to do one thing, the learner finds it easier to do something else. The term 'transfer' also covers the possibility that by learning to do one thing, it becomes less easy to do something else (Thyne, 1966). The potential for either positive or negative 'transfer exists only when the existing cognitive structure influences new cognitive functioning.

In considering teaching the diagnostic thinking process, we are of necessity concerned with transfer of process more than content factors. The discussion must be confined to transfer at its broadest level of generality since, as we have seen in the previous chapter, questions of speciality specificity or even case specificity remain open to debate. This having been said, transfer of learning remains a fundamental consideration for any pedagogic process, and it may be assumed that the one in question concerns not only transfer of substance or content, but also transfer of thinking method or problem solving processes, referred to as transfer of aptitude (Peel, 1967).

Bruner (1960) posed the question, quoted above: How can limited exposure to learning be made to count in the learner's thinking

for the rest of his life? His answer, reflected also in Ausubel et al (1978), indicates the close connection between transfer and structure:

"... the answer to this question lies in giving students an understanding of the fundamental structure of whatever subjects we choose to teach ... The teaching and learning of structure, rather than simply the mastery of facts and techniques, is at the centre of the classic problem of transfer. There are many things that go into learning of this kind, not the least of which are supporting habits and skills that make possible the active use of the materials one has come to understand". (pp 11-12)

Using different terminology, Bruner reflects the dual aspect of transfer: transfer of substance and transfer of aptitude. But in the case of the diagnostic thinking process, what is the substance and what the aptitude? The substance must certainly be, at least, all the basic factual knowledge and information of physiology, anatomy, biochemistry, behavioural science and so on which the clinician may require, together with his knowledge of the possible meaning of all symptoms, signs and clinical information which he may elicit. In addition, the clinician must know what particular tests and investigations are available and advisable for him to use in any particular circumstance. The aptitude must be, at least, his ability to interpret symptoms, signs and clinical information elicited from the patient, as well as the results of tests and investigations, together with his ability to suggest possible diagnoses and, eventually, to select the most likely of these according to pre-determined or specially set criteria, as a basis for definitive action. In all of these, the aptitude transferred is his own thinking processes, however mature or immature, however sophisticated or ill-drawn. To teach with special regard to the diagnostic thinking process may also imply teaching for transfer of self awareness, of the ability to analyse and monitor one's own thinking processes such that, if error of substance be committed, at least error of aptitude will be avoided and even certain types of

error of substance be noted and remedied as a consequence. Such an approach would seem more likely to produce positive transfer since individual differences in cognitive style ("mode of perceptual and cognitive functioning" (Shapson, 1977); "self-consistent and enduring individual differences in cognitive organisation and function" (Ausubel et al, 1978)) express themselves in such problem solving strategies as those under consideration (Gagné, 1965).

Transfer of aptitude may be seen in terms of Bruner's (1973) coding systems. A coding system is a hypothetical construct defined as:

"... a set of contingently related, nonspecific categories. It is a person's manner of grouping and relating information about his world, and it is constantly subject to change and reorganisation".

In this context, transfer of training takes on a meaning more easily related to dynamic cognitive processes which, working in an interactive and unpredictable environment, will never quite reproduce themselves but will be combined and recombined in different ways and subsets according to the demand of the stimulus. This bears some similarity to learning-to-learning transfer. Following Bruner (1973) again:

"You will sense immediately that what I have been describing are examples of transfer of training, so called. But nothing is transferred, really. The organism is learning codes that have narrower or wider applicability".

The codes which we have so far considered are thinking processes applied to clinical information. Yet upon closer examination, it would seem unrealistic to consider the aptitude in isolation from any substance. We may return to Bruner for clarification:

"Much of what has been called transfer of training can be fruitfully considered a case of applying learned coding systems to new events. Positive transfer represents a case where an appropriate coding system is applied to a new array of events

negative transfer being a case either of mis-application of a coding system to new events or the absence of a coding system that may be applied".

The emphasis placed here on the application of coding systems forces us to reconsider the question of transfer of substance. The nature of the substance has been described above but some elucidation of the possible relationship between this and aptitude is necessary. In the clinical, problem solving setting, the transfer of substance must involve using the same, or closely related information in different circumstances, i.e. with different patients or in relation to different diseases. Gagné (1965) refers to this as 'lateral transfer'. It must also involve, for both student and clinician, a learning process which involves building on, perhaps re-interpreting, the substance already learned. Gagné (1965) calls this 'vertical transfer'. More importantly, to be efficient and effective, it must involve storing that substance in such a way that the aptitude coding system can retrieve it appropriately to the flow of information elicited from the patient. We see, therefore, the likely interdependence between substance and aptitude and their mutual mutating effects. In addition, we see the necessity of transfer of this relationship, as well as its individual components from patient to patient, from learning experience to learning experience, laterally and vertically.

The form of transfer being treated with here, and in particular the relationship between substance and aptitude, may be further elucidated by reference to Thyne's (1966) concept of the forceful feature of a stimulus situation. Any stimulus situation, such as a new patient requiring diagnosis, will have a number of characteristics or features, any one of which may evoke a response. But the forceful feature, according to Thyne, is the one, above all others, which actually does evoke a response. The forceful feature, therefore, may or may not actually be the defining feature or cue of that stimulus situation. Thyne

(1966) elaborates on this as follows:

"... we get positive transfer when the forceful feature of the present situation is its cue and when it is also the cue of some established instance of learning. (Transfer is) negative when F, which was in fact the forceful feature, should not have been the forceful feature; that is, when it could be said that the learner had misunderstood the present situation; that is, when F was not the cue (intended) of the present situation. We get negative transfer, then, when the forceful feature of the present situation is not its cue, and when that feature is the cue of some established instance of learning". (p.223)

The complexities of the relationship between transfer of aptitude and transfer of substance are thus clarified. Response to the forceful feature of a stimulus situation and the subsequent thinking processes represents transfer of aptitude. But identification of that feature and its relationship to the actual cue of the situation requires, of course, transfer of substance. Depending upon the manner in which that substance is stored in and retrieved from memory, however, identity of forceful feature and cue becomes more or less hazardous or prone to error. Where different stimulus situations have similar initial presentations, the cue in one may not be the cue in the other, but that same feature may be received as the forceful feature of both. A combination of transfer of substance and aptitude is necessary for resolution of each problem.

It is apparent, then, that the type of transfer which seems appropriate in teaching the diagnostic thinking process could be interpreted in terms of the classical concept of productive thinking. Whereas reproductive thinking is characterised by the solution of problems by means of the existence of stimulus equivalences in the novel situation and in the previously mastered one, in productive thinking past experience is repatterned and restructured to meet current demands (Birch and Rabinowitz, 1951). Although reproductive thinking may appear to be a closer form of transfer, productive thinking is

certainly the type under present consideration if the hypothesis generation and testing model is accepted. The background of past learning and experience is available and essential for restructuring and extension when new situational demands apply. On the other hand, that repertoire of past learning also has the danger of stereotyping thinking and rendering it inappropriately reproductive.

4.1.3 Problem Solving and Learning

Both medical education and clinical practice involve parallel and interwoven experiences of taking clinical histories from patients, solving clinical problems, reinforcing and building upon previous learning either on the basis of clinical experience or by more formal means of education. The outcome of these processes will be their extension: a clinical problem solver who is also continuing to learn. Indeed, such qualities are made explicit in the stated goals of one medical school which has such a document (Royal Free Hospital School of Medicine, 1978). The nature of the discipline makes the relationship between problem solving and learning a central feature, whether for student or practising clinician. We shall see, below, that the practice of medical education is developing in many areas according to this principle.

Gagné (1964, 1965) suggests that problem solving itself is a form of learning. In the following we may substitute 'structure' for 'rule':

"Problem solving may be viewed as a process by which the learner discovers a combination of previously learned rules that he can apply to achieve a solution for a novel problem situation. Problem solving is not simply a matter of application of previously learned rules, however. It is also a process that yields new learning. The learner is placed in a problem situation, or finds himself in one. He recalls previously acquired rules in an attempt to find a "solution". In carrying out such a thinking process, he may try a number of hypotheses and test their applicability. When he finds a particular combination of rules that fit the situation, he has not only "solved the problem" but he has also learned something new". (1965, p.214)

Gagné, essentially, is describing a process of productive thinking, where solutions arise not from blind recall of past experience or blind trial and error, but rather from the requirements of the problem (Osgood, 1953). Gagné (1966) makes the correspondence between problem solving and learning more explicit:

"One of the fundamental criteria of problem solving is that a kind of performance which could not be exhibited before the 'problem' was solved can be exhibited after the 'problem' is solved. In other words, the observed events in problem solving comprise a change in human performance, and this in turn leads us to infer a change in human capability".

Ausubel et al (1978) relate problem solving back to the concept of structure and transfer already discussed. The solution of any given problem involves reorganisation of past experience to fit the current demands, and:

"Since ideas in cognitive structure constitute the raw material of problem solving, whatever transfer occurs, positive or negative, obviously reflects the nature and influence of cognitive structure variables".

The identity of such variables may become modified as successive solutions are based upon cognitive reorganisation.

On the basis of the discussion so far, it would seem reasonable to conclude that problem solving skills could best be learned by problem solving learning. Indeed, evidence is available that by solving problems we develop 'learning sets' and acquire problem solving skills which may be applied to future problems (Manis, 1968). However, such work as this has normally considered only well defined problems and does not account for the many possible errors of aptitude that may occur when attempting to apply an old solution method to a new problem. A more rational and generalisable approach is described by Peel (1967) based on the premise that "whatever the problem, certain basic thought processes seem to be involved - the imagining of possible explanations, the selection of the correct ones, the capacity to relate explanations

with the data of the problem in a logical way by inference and deduction". The two attempts at teaching these general skills which Peel reports have proved encouraging. However, Peel also suggests that different subjects "seem to emphasise different aspects of concept formation and usage and so may entail different elements of thinking". The generalised problem solving skills defined seem analogous to the hypothesis generation and testing model and to that extent it would seem reasonable to teach for such thinking skills as these. However, if Peel is correct in suggesting subject differences, then the particularities of clinical problem solving thinking may remain to be defined before a special programme of teaching can be undertaken. However, this is not to suggest that specific thinking skills should be taught and learned. The diagnostic thinking process does not seem to be of the same order as, for example, long division or the two cord problem, where the salient features (rules about the permissible forms of numerical operation and perception of properties of objects, respectively) can be defined quite clearly. On the contrary, the diagnostic thinking process must rely heavily upon each individual's own cognitive style and propensity. Externally imposed forms, such as method of data collection or recording, may have no effect upon the problem solver's actual thinking processes. To specify the particular properties of the clinical problem solving process, therefore, does not necessarily imply specification of a single or particular form, but may mean only the identification of a range of possible processes.

From a theoretical point of view, then, this latter option seems the more tenable, particularly if we consider the two types of learning that may be important in problem solving as identified by Birch and Rabinowitz (1951). The first type involves the acquisition by the problem solver of "certain broad, nonspecific, general notions about the properties of the object or method experienced". This type "seems to provide the repertoire of experience essential for productive

thinking". The second type of learning involves "the acquisition of experiences which convert the initial perception of broad, general properties of an object into perceptions of limited functional characteristics". Such a type of learning "limits the range of perceptual organisations capable of being developed by the subject and so interferes with problem solving".

4.1.4 Conclusions

We have considered three separate yet closely interrelated aspects of teaching and learning and have described their relevance and importance for development of the diagnostic thinking process in student and practising clinician. It is against these three criteria that the subsequent discussion of various teaching methods already devised will be measured. The discussion so far leads us to a number of conclusions:

Firstly, structural aspects of the task, the teacher and the learner must be taken into consideration. In addition, appreciation of those structural aspects must be one aim of the teaching and learning process. Getzels (1964) discusses the distinction between "ugly thinking" (problem solving based on trial and error) and "productive thinking" (solutions based on cognitive reorganisation) and concludes that:

"What is especially significant for instruction from this point of view is that the person 'naturally' seeks the inner structure of a problem situation if he is unfettered by habits developed through training in blind association and 'ugly' trial and error learning". (p.244)

The hypothesis generation and testing model implies exactly such a search for inner structure. It is to the nature of this search and its successful and unsuccessful forms which the pedagogy of the diagnostic thinking process must address itself.

Secondly, provision must be made for transfer of learning of both substance and aptitude. Such transfer also pre-supposes

learning appropriate structures for both substance and aptitude such that each may be retrieved, accessed or reorganised according to the demands of a variety of tasks. Therefore questions of the manner, nature and structure of knowledge and skill acquisition must be considered.

Thirdly, the nature of problem solving as a form of learning itself must be taken into account. To this extent, it is not possible to 'practise' problem solving without learning or assimilating the structure of that practice. We would conclude, therefore, that although problem solving skills must ultimately be learned by problem solving learning, this must be of a special type. Yonke (1979) advocates teaching the diagnostic process as a separate cognitive process.

To return to our original theme, we would agree and suggest that the structural aspects of the problem solving process must be teased out and presented to enable self analysis and self awareness in the dynamic process. Goldberg (1968) demonstrates the lack of effect of non-analytical problem solving learning. However, given an analytical framework, problem solving learning becomes a productive process in its own right. Piaget (1970), in his own terms, reflects this principle:

"... knowledge is derived from action, not in the sense of simple associative response, but in the much deeper sense of the assimilation of reality into the necessary and general co-ordinations of action. To know an object is to act upon it and to transform it, in order to grasp the mechanisms of that transformation as they function in connection with the transformative actions themselves. ... it follows from this that intelligence, at all levels, is an assimilation of the datum into structures of transformations, from the structures of elementary actions to the higher operational structures, and that these structurations consist in an organisation of reality, whether in act or thought, and not in simply making a copy of it". (p.29).

Piaget also adds "the heartbreaking difficulty in pedagogy" that the best teaching methods - active methods - are also the most difficult. In learning the diagnostic thinking

process, active methods cannot simply mean clerking patient after patient after patient, discussing findings and conclusions during the ward round or tutorial, and divining the cognitive error or the nature of correct thinking. The difficulty in active learning, here, must be to attain the correct balance and relationship between theory and practice, between understanding the structure of one's own thinking processes and the structure of the substance, and using these structures to solve clinical problems and thereby, perhaps, altering them.

4.2 Approaches in Teaching the Diagnostic Thinking Process

Barrows et al (1978) voice the reasonable anxiety that:

"... the usual subject oriented education in medical schools does not require actively practised reasoning skills as the students learn. There is concern that the traditional approaches to medical education may actually diminish the natural problem solving skills possessed by students before they enter medical school". (p.16)

The authors trace this development, and the associated division between the basic sciences and the clinical courses, back to Flexner's (1925) study of American medical schools. Elstein et al (1978) take a rather harder line, leaving the reader in little doubt about the direction of their arrows:

"Gaps in medical knowledge are easy to spot when they produce major obstacles to problem solving, and remedies are equally obvious. In contrast, failures to use efficient strategies are difficult to discern and even more difficult to remedy. It is no accident that medical schools traditionally have packed the curriculum with as much factual content as time and student capacities permitted, and have offered little or no formal guidance in problem solving strategies". (p.271)

Perhaps the criticism is a little strident and should be tempered by recognition of our still shallow knowledge of the clinical problem solving process. But the over predominant base of factual knowledge and the failure to teach use of that knowledge for clinical problem solving remains a criticism

echoed by other workers (Meals, 1973; Hiss and Pierce, 1974; Barrows and Bennett, 1972; Barrows and Mitchell, 1975; Barrows, 1976; Yonke, 1979; Maddison, 1978; Miller 1978; Wyn Pugh et al 1975; Lloyd et al, 1976). Sherwood (1978) speaks of the tyranny of information gathering to the exclusion of teaching what to do with it, while Iansek and Balla (1979) point out that many teachers remain in ignorance of their own diagnostic thinking processes.

Despite such criticisms as these, many curriculum changes are under way, and a variety of teaching strategies have been devised to assist the student in developing his diagnostic problem solving skills. These will be considered under five headings: the integrated curriculum; problem based learning; training in hypothesis generation; cognitive skills training; and, computer assisted training.

4.2.1 The Integrated Curriculum

Stenhouse (1975) discusses Bernstein's (1971) distinction between collection and integrated curricula. In a collection type, the contents are clearly bounded and insulated from each other; in an integrated type, the contents stand in open relation to each other. Armstrong (1977) suggests that the pre-clinical curriculum, comprising a study of the separate basic medical sciences, constitutes a collection type curriculum, whereas the clinical curriculum which follows it is of the integrated type where "although learning may be mediated by specialities, there is little insularity from one speciality to another and knowledge gained in one can be used in another". However, if we follow this argument through, and consider the relationship between pre-clinical and clinical curricula, it is clear that these are not, in their traditional form, integrated. From the point of view of structure, transfer and problem solving, as discussed above, the undergraduate medical curriculum as a whole may be criticised. The structure of biochemistry as taught in the pre-clinical course, for example, may well not transfer to the study of endocrine disorders in

the clinical years. The academic sciences of the pre-clinical years, set outside the clinical problem solving context of the clinical curriculum may well require considerable cognitive restructuring or reorganisation before they assist the diagnostic thinking process in an efficient and effective manner, for disease does not differentiate between separate scientific territories or levels of enquiry into bodily structure and function. The collection curriculum also is implicitly questioned in Ausubel et al's (1978) argument that:

"It is less difficult for human beings to grasp the differentiated aspects of a previously learned, more inclusive whole than to formulate the inclusive whole from its previously learned differentiated parts". (p.190)

Especially so if structural aspects of those parts make transfer to the inclusive whole difficult in the initial instance.

Awareness of this type of contradiction between structure of learning and method of use of information has led to an upsurge in attempts to integrate both within and between the pre-clinical and clinical courses and thus facilitate the necessary criterion of transfer of learning.

Miller et al (1961) considered at that time that "at the medical school level probably Western Reserve alone has made a concerted effort to build an entire curriculum around the principle of transference". This American medical school - Case Western Reserve - was the first to develop a co-ordinated teaching programme, the establishment of which began in 1945 (Spilman, 1965). The integrated teaching was subject based and inter-departmental with the collaborative effort of faculty members from different departments and disciplines. The curriculum included a project teaching programme and deliberate correspondence between pre-clinical and clinical material, always with the maximum participation of the student.

Since that time, forms of integration have been adopted in many British medical schools. The General Medical Council (1977)

showed that six of the 38 British medical schools have adopted or planned curricula in which vertical and horizontal integration (pre-clinical - clinical and between disciplines) is a major goal, while some element of inter-disciplinary teaching is found in all medical schools. "There is increased recognition that teachers have a common goal and that it may be advantageous to work together to attain it" (Harden et al, 1978). In 1973 (Association for the Study of Medical Education, 1973) alone, the Newcastle upon Tyne Medical School gave a retrospective appraisal of its ten years of a curriculum in which the pre-clinical years comprised both temporal co-ordination and subject integration and the clinical years constituted 'true multi-disciplinary teaching' based on courses concerned with bodily systems rather than separate branches of medicine or surgery. At the same conference the Nottingham Medical School reported on its vertically and horizontally integrated programme for the first three years of the curriculum, based on three thematic courses (The Cell, Man, The Community). As the course evolved, so did considerable merging of pre-clinical and clinical teaching (Olson, 1976). Dundee Medical School also reported on its examples of temporal co-ordination and subject integration.

In 1971, the Southampton Medical School took its first students into a curriculum based on bodily systems courses, with patient contact for the student from the beginning (Acheson, 1976; Elstein and Forbes, 1976). It is argued (Howell, 1976) that the hazards of the traditional curriculum may therefore be avoided. In the traditional curriculum:

"The student himself is expected to undertake the horizontal integration of such courses and to develop an understanding of the different systems of the body in health and disease".

Vertical integration in the Southampton curriculum arose from the view that:

"... 'clinical relevance' is important in the early stages of the medical curriculum both as a stimulus to learning and as illustrating the emphasis to be

placed on items of factual information ... The medical student proceeding to clinical medicine needs integrated models in his everyday clinical practice, and a programme of learning which embraces the disordered at the same time as the normal and the abnormal and ~~permits~~ economies in learning and teaching".

Although the student's acquisition of knowledge and skills is still removed from the clinical situation in which he will have to use them, the Southampton curriculum obviously represents a concerted effort to deal with the structural aspects of learning and the problem of transfer. One cannot divine how successful that attempt will be in terms of the future development of clinical problem solving performance in students, but it is, nonetheless, being made.

It is clear that the concept and implementation of the integrated curriculum is far from monolithic. It ranges from temporal juxtaposition of courses to the development of systems based teaching, from horizontal integration to the vertical integration attempt to cross the traditional Rubicon between pre-clinical and clinical teaching. But how far do these strides forward go towards attaining the three criteria of structural integrity, transfer of learning and development of problem solving skill?

This question can only be answered speculatively; but first it might be useful to refer back to Bernstein (1971). Integration, as he uses it, refers to the subordination of previously insulated subjects or courses to some relational idea, some idea such as a bodily system. Integration can thus be handled either by a single teacher or a group of teachers. If the latter, then surrender of traditional sovereignties is necessary. Stenhouse (1975) points out that if the relational idea which must predominate in those involved in teaching is not sufficiently powerful, then the integrated study may not integrate. In the case of temporal co-ordination of topics, this is almost certainly so. Integration will still depend upon the learner's ability for cognitive reorganisation or

restructuring.

To return to our three criterion; our first general assessment must be that the integrated courses and curricula discussed do not intend to treat with the diagnostic thinking process per se, but rather to structure knowledge such that its retrieval in the process of clinical problem solving is facilitated better than learning on the basis of scientific, medical or surgical speciality would allow. In other words, integration attempts to cope with structural and transfer aspects of learning, but not with problem solving skills. In psychological terms, it is not possible to judge whether information learned in an integrated curriculum is structurally more suitable to clinical practice or the diagnostic thinking process than information learned in a traditional curriculum would be. It may be that one requires less structural reorganisation than the other, but without further data one can only refer to common sense argument.

Considering the matter from the point of view of later teaching during the clinical course, it would seem reasonable that the substance of the integrated curriculum is structurally more appropriate than the substance of the traditional, speciality based curriculum. This conclusion can only be reached, at present, on the basis of the structure and content of the clinical history which is, itself, largely systems based. However, whether or not the substance learned in an integrated curriculum gives rise to, for example, a more precise or wider ranging interpretation of clinical information elicited from the patient is not known.

In conclusion, the integrated curriculum would seem to have the potential for building on structural and transfer aspects of learning, but its likely effect on the learner's development as a clinical problem solver remains obscure. However, in that appropriate structures and transfer of learning are necessary pre-requisites for effective clinical problem solving, we may

conclude that the integrated curriculum, in particular if vertical integration is a strong feature, is likely to have a positive, mediated effect upon the development of the diagnostic thinking process.

4.2.2. Problem Based Learning

"When discussing innovations in medical education, one can easily get the feeling that problem solving as an educational technique is a new thing. But problem solving has been used as an educational technique in medical education for many years, although during the last period of time the meaning of the term has changed. Nowadays, the term is used not only to describe how the symptoms and signs are put together to make a diagnosis, but also to imply an awareness of the way medical students are taught to use these stages and techniques". (Ström and Walton, 1978)

So concluded a discussion group at the 1977 Conference of the Association for Medical Education in Europe on 'Innovations in Medical Education'. Despite the contention, only four medical schools yet boast curricula based upon a problem solving approach (McMaster, Shiraz, Maastricht, Newcastle NSW), but many courses include problem solving exercises of some type.

The most recent curriculum to be based upon a problem solving approach is that of the new medical school in Newcastle, New South Wales. This curriculum;

"... is based on learning through the management and solution of clinical problems ... The study of these problems will ensure that student learning is relevant to contemporary medical practice. ... (The) student will use the clinical problem ... to identify and acquire the knowledge, understanding, skills and attitudes he needs to manage (the) patient". (Clarke, 1978).

In dealing with each clinical problem, the student will also learn the relevant biochemistry, physiology, anatomy, pathology, risk factors, epidemiology and prevention, physical examination, investigations, management of the patient, and other family and social aspects. In this way:

"The student will integrate all these in the context of the patient, and learning will be enhanced by the motivation which will result from the clear relevance of the task. Thus the basic sciences will not be learned as separate disciplines, but will be studied and applied in a clinical context from the first to the last day of the course".

Problem based learning in Newcastle is thus seen as having three major advantages; firstly:

"The ability to view every encounter with a patient as a problem enables the doctor to determine what is new and unknown to him, where to obtain the additional information he needs and how to apply it to the problem, thus adding actively to his store of knowledge, understanding and skill". (Engel and Clarke, 1979)

Secondly:

"As the student is encouraged to apply his newly gained learning he reinforces this learning and obtains rapid information on the success of his studies". (ibid.)

Thirdly:

"... problem based learning allows complete and continuous integration of the basic and clinical sciences. We know of no successful method for achieving this, other than through learning in order to deal with problems". (ibid.)

It would appear from this that structural and problem solving aspects of the learning task are being addressed by the Newcastle curriculum, although we may wonder whether practice at solving problems as such, without specific feedback on the thinking process rather than the knowledge base, will actually encourage appropriate problem solving thinking or an awareness of one's own thinking processes. This would seem doubtful. Shulman and Elstein (1975) point out that "the essence of learning is not merely doing, but thinking about what one is doing". To encourage students to take a problem solving approach, and to perceive clinical problems and their solution as the central feature of medical education and clinical

practice seems, on a commonsense level, to be a reasonable strategy. To attempt to create the habit of continuing education by constant encouragement to formulate and answer searching questions (Engel and Clarke, 1979) surely is also an important aspect of undergraduate medical education.

However, despite the undoubted horizontal and vertical integration of information which problem based learning achieves, its actual strengths may, possibly, rest there. Let us firstly consider the question of structure. As we have already discussed, the structural properties of an individual's concepts and knowledge are important because they determine how he will acquire and use knowledge in the future. But what are the structural properties of knowledge acquired through problem based learning? Efficient and effective storage of information in memory requires considerable organisation (Hunter, 1964). What are the organising agents in problem based learning? These questions are important, simply because the senior student and the practising clinician do require reference to material stored in memory. Clinical problem solving is not simply a matter of formulating and answering searching questions: substance must complement aptitude. How is that substance to be stored to make it most easily and appropriately retrieved and related to other pieces of knowledge and experiences? The difficulty seems to arise from the view of problem based learning as a teaching method to achieve integration and build knowledge appropriate to clinical practice. Learning through solving problems, however, is not quite the same thing as solving problems and, incidentally, learning. The question of structure, therefore, remains open and interesting. Complementary to this is the question of transfer of learning. Efficient transfer depends upon appropriate structure. How is learning set in the context of one problem to be transferred to a new problem? Finally, it has been indicated already that learning through problem solving is not necessarily learning to solve problems or

acquiring insight into and the means to evaluate one's own problem solving strategies. Without some special input, problem based learning does not necessarily achieve these ends.

Having been thus critical, it is necessary to suggest what might be a correct perspective. If questions of structure, transfer and problem solving remain unclear in Newcastle, NSW, questions of integration, orientation, the centrality of patient and clinical problem and attitude towards learning seem to be elegantly resolved.

The McMaster curriculum has addressed some of the questions which we have so far failed to answer. Barrows (1973) discusses the educational rationale for problem based learning:

"In problem solving the active participation by the learner in his learning and the development of inquiry or problem solving skills along with self directed study techniques, makes this method attractive ... In problem based learning the student must observe, think, define, study, analyse, synthesise and evaluate. These active and self directed processes are necessary to understand, define, solve or manage any problem".

Specifically, Barrows and Sibley (1978) cite advantages similar to those discussed in relation to the Newcastle, NSW, curriculum, but Barrows (1973) also addresses the questions of structure and transfer:

"With problem based learning there is evidence that information is acquired in a set around the problem under study. This provides for mental associations that allow the student to retrieve the information he has acquired when working with other problems that may require the same information. These information sets, formed around a particular problem, allow for succeeding problems in a particular area, such as neurology, to be solved with increasing speed and facility. As the student acquires an increasingly rich number of learning sets about different problems and uses these in the solution of new problems ... he becomes increasingly more able to solve complex and novel problems. The information the

student acquires is not only more easily transferable to new situations and problems but the acquired information may be better remembered".

No evidence is produced to support these arguments, other than acquisition of information as a set around the clinical problem. Details of variety of problem presentation, identity of the forceful feature, and ability to solve new problems by reference to stored information would all assist in resolving the problems of transfer and structure. At McMaster, problem solving aspects are also considered:

"It is important to identify the student's progress in developing problem solving skills. Are his problems in problem solving identified, are they rectified by the student? Does the student go to enough pains on his own to determine whether his problem solving, approach, references, and self study skills and quality of the information he has gained are appropriate?"

To this end, self awareness questions are included in the problems (Learning Resources Design Project, 1975). Learning and problem solving at McMaster appear to hold equal pride of place. This may be illustrated with reference to the McMaster 'problem boxes' (Barrows and Mitchell, 1975).

The problem box contains a printed clinical problem manual and related study materials in various media. The problem is presented as it would unfold to the clinician. For example, after "A 55 year old man complains of a headache", the manual stops and the student must write down his thoughts about this. A few more sentences of the problem then unfold on the next page, and so on. The questions actually asked of the patient by the original examiner are reported, and the student is asked to comment on them. As the case evolves, the student can compare his approach with that actually used. A separate manual leads the student into a study of all aspects of the case - physiology, anatomy, clinical techniques, etc. A list of learning resources is also provided, and faculty are always available to discuss the problem. With some boxes the student may deal even more realistically with the problem by

interviewing and examining a simulated patient trained to simulate the patient presented in the problem box. Problem boxes were used in conjunction with an 'educational prescription' tailored to the learning needs and style of each student individually.

Following on from the problem boxes, has been the development of the Portable Patient Problem Pack (P4) (Barrows and Tamblyn, 1977), which was designed specifically "to develop the student's problem solving or diagnostic skills in a manner consistent with the skills of the practising clinician".

A scoring system is incorporated to assess the student's clinical reasoning skills and the cost of his evaluation of the patient. A P4 unit, simulating one patient problem, consists of 280 coloured cards categorised into types of action that can be taken (white for questions on history, blue for items of physical examination, etc.). On the front of each card, below the title, is a series of questions the student should ask himself before taking action of that type. Thus P4 includes a component designed "to help shape the effectiveness and efficiency of the student's problem solving or clinical reasoning skills". The back of each card gives responses to the action indicated on the front. This may include referral to slides, X rays, laboratory data, etc. Some cards are designed to facilitate the student's interaction with the problem and evaluation of his own performance. The student works with P4 by spreading the title cards in front of him, reading a card to acquaint himself with the patient's problem, and selecting cards until he has handled the problem as far as he can or wishes to. This process may include time for reading, study and conference with faculty.

Problem boxes and the P4 format, then, represent teaching methods designed specifically to improve students' clinical problem solving skills.

Such simulations as these developed from the original work of

Rimoldi who developed Diagnostic Management Problems during his studies of the problem solving process (Rimoldi, 1955; 1961). Rimoldi's method involved the use of an array of cards listing all possible tests and questions, from which the subject must select in response to being presented with information about a 'patient'. By looking at the questions asked and their order, an evaluation of the subject's thinking processes was attempted. The approach assumes that at every successive step the problem changes and that what the subject knows and what he may still want to know are not fixed properties of the problem, but vary as the solution develops. Using this method, it is also possible to identify the questions which the subject omits thereby defining the areas which he may consider to be redundant or irrelevant.

Based on Rimoldi's (1961) work, Helfer and Slater (1971) devised problems and computer calculations to give students feedback on their own process and outcome performance. Each problem was presented in 96 cards consecutively numbered. A specific historical fact, physical finding or laboratory result was given on each card. The student was first told the setting of the problem, such as 'Outpatients' Clinic', given a brief abstract of the case, and provided with an index sheet which itemised the type of information on each card, such as 'Skin', 'Blood Pressure' or 'Skull X ray'. The student selects cards in whichever order he prefers, and records his selection. The primary and secondary diagnoses, and the point at which the primary diagnosis was made, are also recorded. The computer output provides feedback to the student on: the diagnostic score to show accuracy of the diagnosis made and appropriateness of the point at which it was made; the process score, derived by comparing the student's sequence with that of experts; the efficiency score derived from the total number of helpful cards selected divided by the total number of cards chosen; cards not selected in the appropriate sequence; cards considered helpful by experts but not selected by the student; cards the student failed to

select before he made his diagnosis which experts felt should have been selected; unhelpful cards selected; and harmful cards selected.

Such a method as this, then, provides the student with ample information for him to compare his selection and processing of information, as compared with that of an expert. But such a teaching method is, of necessity, very much based in teaching the significance of factual information, and interpretative value of clinical and laboratory data. Such a method, however, should also prompt the student to think more clearly and consciously about the nature of his own information gathering strategies. But the generalisability of those strategies, and the differentiation between parts which are attributable to the case and parts which are attributable to his own problem solving thinking processes are likely to remain obscured.

Rimoldi's diagnostic management problem technique was adapted, firstly by McGuire and her colleagues at the University of Illinois College of Medicine, and the resultant simulations were called patient management problems. A book of examples has been published (McGuire and Solomon, 1971) and many adaptations have been reported (Newble, 1975). Patient management problems have been used for assessment as well as teaching. Each problem begins with some information about the patient, including the main complaint. The problem solver must then gain more information for diagnosis and management. A list of further available information is given, and figures, slides and X rays may be used to provide some answers. The student thus collects data in a sequential manner, either linearly (e.g. Hubbard, 1971) or in a branching way (McGuire and Solomon, 1971). A variety of technical devices is available to obscure answers before selection; these include paper tabs, invisible ink and compounds that can be erased. A record of selections is kept. Cuing is reduced by offering a large number of options, and the format facilitates observation of sequential decision making based on feedback from the answer sheet. McCarthy and Gonella (1967) used a similar approach

for evaluating and teaching clinical competence, but maintain that the presence of even long lists of possible relevant choices provides a cuing effect which is not present in actual clinical situations. This effect is most marked for poorer students (McCarthy, 1966). An additional problem is that the cumulative nature of data collection may also cause cumulative error (Berner et al, 1974). In a factor-analytic study, Juul et al (1979) concluded that patient management problems measured data gathering and management components of medicine problem solving.

Although patient management problems have been widely used, they are subject to the same criticisms as the diagnostic management problems. They are primarily knowledge based, providing sophisticated exercise in the collection and sequential use of medical information. There is little evidence that such a teaching method will enhance the cognitive problem solving process of the learner in a real clinical situation or that it provides him with any insight into his own thinking processes and use of information. Shulman and Elstein (1975) quote findings that patient management problems provide indication of inadequate clinical performance, but they do not indicate what a physician will actually do in practice. This may indicate that learning by patient management problems does not effect generalisable problem solving skills. In other words, problems of structure, transfer and problem solving remain unresolved.

In conclusion, a variety of different approaches in problem based learning have been discussed and evaluated from the point of view of structural, transfer and problem solving aspects of education. It appears that problem based learning as such need not enhance any of these particular characteristics, although enhancing others equally as desirable but not related to the development of the diagnostic thinking process. We may leave the final word to Luchins and Luchins (1950) who were speaking of mathematics, but whose lesson remains pertinent:

"One implication for education is that in teaching mathematics it is not sufficient to make the problems more concrete, more life like. The trend toward concretizing mathematical problems by relating them more closely to everyday activities is in part motivated by the desire to make the subject matter more meaningful to the child; but this need not result in giving the child a better insight into mathematics - he may still repeat blindly certain rules and formulas. What are needed are teaching methods which will lead to understanding of the structural qualities of mathematical concepts and encourage productive thinking".

4.2.3 Training in Hypothesis Generation

Since hypothesis generation and testing has been the dominant cognitive model of the diagnostic thinking process, some workers have attempted to train students, and clinicians, to generate better, more accurate, more appropriate diagnostic hypotheses. For example, Elstein et al (1978) report a study designed to develop and test a procedure for training students to generate diagnostic problem formulations based on cues elicited during the first four to six minutes of the clinical interview. The method of training involved the use of specially prepared films of the initial segment of the clinical encounter, from the physician's eye view. The student viewed the film and was given basic demographic information about the patient. He gave impressions and diagnostic problem formulations, filled out a set of response sheets and wrote a brief tentative assessment. Feedback materials based on physician performance were provided; these materials concerned either process or outcome or both. Results of this process showed that the training model was effective in improving second year students' skill in generating a set of initial problem formulations. In addition, the training model was just as effective when providing outcome feedback only, as it was when outcome and process feedback were given. This latter finding is interesting, possibly implying that practice is primary in developing the skill. The work

represents an attempt to isolate one part of the diagnostic thinking process and train for that in particular. It seems successful in this attempt, although no explanatory theory accounting for this success and the satisfactory influence of outcome-only feedback is given.

It will be noted that the approach just described is problem oriented. The problem oriented medical record has been presented as a teaching tool (Weed, 1968), but does not pretend to enhance the cognitive processes of the physician as such; rather, it assists the problem solver to systematise clinical experience and use and communicate clinical information by means of the clinical record (Lloyd et al, 1976). A side effect of this, of the formulation of problem lists in particular, may be to encourage efficient and effective possible diagnostic formulations, but research results are not reported.

An attempt to encourage students to formulate early diagnostic hypotheses is also reported by Taylor et al (1978) in gynaecology, in response to the Michigan and McMaster work. The course consisted of core document objectives, flow charts, small group sessions designed to encourage appreciation of the role of hypothesis formation, simulated patient management problems and a problem solving approach. The results of the course were not entirely satisfactory, only 73 per cent of students learning to form and revise hypotheses adequately. However, two factors may account for this: firstly, it is not clear to what extent students were explicitly made aware of this process; and secondly, the course was designed to encourage the student to make "broad, general diagnostic hypotheses rather than the narrow hypotheses typical of beginning students". As discussed above (Chapter Three, Section 1.2), this view of Barrows and Bennett (1972) is now open to some doubt, and so it may be that Taylor et al's (1978) teaching methods were running counter to the cognitive inclinations, and effectiveness, of the students.

Elstein et al (1978) devised a set of five heuristics to "improve the problem solving performance of advanced medical students functioning in a hypothesis guided mode". Students were thus encouraged to think about the relationship between their hypotheses and the information available, between the hypotheses themselves, the effect of new hypotheses in causing re-interpretation of old information and so on. Results did not strongly support the research hypothesis of improved problem solving performance due to heuristic training. It is possible that the heuristics themselves were too specific and merely added a confusing, complex dimension to the thinking process rather than encouraging self awareness.

The P4 units, described above, are introduced to students in the context of hypothesis generation, with the implicit assumption that many hypotheses should occur to the student (Learning Resources Design Project, 1975):

"If your definition of the patient problem only produces one hypothesis initially then use it alone and direct your questions and examinations of the patient towards establishing or denying its validity. If you eventually end in a blind alley after following what you feel are appropriate investigations, you can then (1) reconsider your problem definition; (2) ask more general questions of the patient; (3) survey with system inquiries for other symptoms or complaints; or (4) perform other items of examination until new data suggest a revision of your crystallization or new hypotheses are suggested and then proceed in a similar manner".

The guide goes on to explain the importance and use of diagnostic hypotheses in terms of diagnosing, dealing with information arrays, memory and so on. In addition, students are reminded to be alert for items of information which do not fit into any of the hypotheses generated. Strong advice is given that the first hypotheses to be generated during any clinical encounter should be broad, such as "something wrong in the left hemisphere", and more specific ideas should only be formed as the data unfolds. Finally, students are warned against premature acceptance of any easy or obvious answer.

Such an approach to teaching should encourage transfer of aptitude from case to case, and seems eminently promising, given its context in problem based learning as described in the previous section. The only point of doubt immediately obvious is the unwarranted, although not extreme, advice to generate broad hypotheses before becoming more specific. Such a tactic is not necessarily appropriate in all cases, since hypotheses can only arise from the information given and the associated structure of information in memory. It is possible, also, for a broad hypothesis to be as inappropriate as more specific ones. Attempts to follow this advice where the information and structures suggest specific ideas may be more confusing than clarifying. However, awareness of the advantages and dangers of the process are a positive contribution to training in clinical problem solving.

Training in hypothesis generation has been attempted in other fields. Klein et al (1969) studied the role of anxiety in learning a task which required subjects to formulate original responses to complex stimulus situations. Feedback was given to subjects in the form of a list of "acceptable" hypotheses. Attempts were made to improve both the quality and quantity of hypotheses, but only the latter actually showed improvement. Dienes and Jeeves (1965), considering the structural derivation of hypotheses, believe that false ones should be followed up since "hypotheses can only be rejected by recourse to the actual situation in which they were relevant". This is the substance, not aptitude, aspect of hypothesis generation and may prove a fruitful approach.

In conclusion, from the point of view of structure, transfer and problem solving, such training in hypothesis generation as we have discussed may well assist structuring operations and help the learner to appreciate how he relates his own knowledge structure to the presenting stimulus situation. An appreciation of aptitude or problem solving process may transfer from case to case directly since an operation which could be

identified as hypothesising will certainly occur in every instance. Awareness of the dangers of this mechanism is also to be encouraged. In this way, problem solving itself should be enhanced. Training of this type would seem, at the moment, to be limited only by the severely limiting description of the diagnostic thinking process as 'hypothesis generation and testing' and concomitant errors such as the assigned temporal relationship between broad and specific hypotheses. The general approach, however, seems worth building upon and refining.

4.2.4 Cognitive Skills Training

Another approach in medical education, although not very widely used, has been to train students in general problem solving skills and awareness of their own cognitive processes, with the belief that the results of such training would generalise to their clinical problem solving task.

Apart from one unsuccessful attempt to use feedback on direct observation of the student - patient encounter (Hinz, 1966) and a suggested use of decision trees for teaching management of uncertainty (Knaggs et al, 1974), the work of Abercrombie (1960) and related work of Lavelle (1977a; 1977b; 1978) represent this systematic approach to teaching clinical problem solving thinking. Abercrombie (1960) devised and implemented a course for medical students of eight 90 minute discussions in groups of twelve students. The content of the discussion sections was seeing/perception, language, classification, evaluation of evidence and causation. Results showed that students who had taken the course did statistically better than others in four respects - they tended to discriminate better between facts and conclusions, to draw fewer false conclusions, to consider more than one solution to a problem and to be less adversely influenced in their approach to a problem by their experience of a preceding one. That is, overall, they were more objective and more flexible in their behaviour. The aim of the course - "to make it possible for the student to relinquish

the security of thinking in well defined, given channels and to find a new kind of stability based on the recognition and acceptance of ambiguity, uncertainty and open choice" - seems to have been fulfilled. This work has since been followed up (Abercrombie, 1978).

Lavelle (1977a; 1977b; 1978) also takes the approach that to train for general cognitive skills may specifically apply to the clinical situation and that by becoming conscious of factors which unconsciously influence judgments, more valid judgments may be made. The operations examined by Lavelle (1977b) have been: abstracting information from the literature; perceiving a number of objects; observing an object; describing a process; accessing unstructured (sic) memory by association; forming hypotheses; solving problems; and communicating with others. The parts of the course described in Lavelle (1977a) are specifically related to clinical topics and situations. Evaluation results of the courses are not yet available, although some improvement in information processing skills is reported (Lavelle, 1977b). However, "whether the improvement is sustained, or is applicable to medical decision making remains to be established".

Evaluation of such work as this against our three specified criteria is difficult. One can do little more than conjecture. Perhaps awareness of one's own thinking processes and cognitive bias may transfer to any substance or occasion. Abercrombie's work suggests this as a possibility, but the indirect nature of the method also suggests problems of structure and transfer.

4.2.5 Computer-Assisted Training

Although still not very widespread, the use of computer based systems for training in decision making is sufficiently ample to merit some discussion. De Dombal (1979) reports such programmes in England, France, Scotland, Canada and the United States.

Interactive computer programmes have been used for sequential teaching similar to that of the patient management problems. It is suggested that computer simulations offer the opportunity to use expert data in the calibration of student error (Johnson et al, 1975) - an advantage shared by the card packs and paper simulations already described - and may be used to avoid the cuing effects of the patient management problems (Berner et al, 1974), although problem boxes can avoid this effect also. De Dombal and his colleagues at the Leeds University Medical School have been foremost in Britain in developing computer assisted learning systems, and have compared these with similar non-computerised systems.

De Dombal et al (1969) described the initial Leeds computer assisted system for learning clinical diagnosis. The student obtains information by entering questions to the computer via a teletype and receiving responses immediately on the printout. The information is stored in the computer under headings or files, and, in that the student is not allowed to leave a file until he has collected all the useful information it can give, there is some control over the sequence in which he gathers data. At any point in the process, the student can refer to a 'meaning', 'consultancy' or 'medical teaching' file to assist him in understanding terminology, interpreting symptoms and signs or taking a history and physical examination. A special 'help' file is also available if he feels that he is still making no progress. Having collected enough data, the student makes his diagnosis and this, together with the way in which he gathered that data, is evaluated by the computer. This is then discussed and reviewed with the teacher.

De Dombal et al (1969) discuss the problems of such a simulation, especially in terms of its unreality and the interference caused to the student's thinking processes by having to type in questions, and these problems were gradually to be resolved. However, the major advantage of the system was seen as its ability to ensure that "each diagnostic procedure is carried out logically and orderly" and to give immediate monitoring

and correction of errors in techniques. It would seem, then, that this system encourages the student to learn a certain rigour of approach in eliciting information, but deals only with the knowledge base and not with his manner of combining information. Likewise, the request for a diagnosis at the end of data collection obscures the dynamic thinking processes and reasoning of the student. More recent approaches, however, have overcome some of these drawbacks.

In 1972 were reported the results of an evaluation of four different modes of diagnostic simulation, including a computerised mode (de Dombal et al, 1972b). The four modes evaluated were: MINISIM based on the diagnostic management problem card format; VERBAL in which an operator kept the cards and the student interrogated the 'patient' operator who responded by giving the appropriate card; VERBAL + CAD in which the operator answered the student's questions instead of giving cards. In addition, students could submit their case history for Bayesian analysis and ODSAL (off-line diagnostic simulation for additional learning) involving a teletype in an off-line format, with a paper-tape punch and reader. This programme, in a set form, stopped at intervals for the student to type in his differential diagnosis. The tape finally gives the results of the computer's analysis of the case and the (real life) operative diagnosis and outcome.

Results showed no statistically significant differences in short term retention of information between students who had used a simulator and those who had not. However, there were statistically significant differences between these two groups on tests of information collection and evaluation. No statistically significant differences were found between the four modes of simulation. De Dombal et al (1974) report further modifications to the system - in particular, ODSAL gained an on-line facility and a comparison of students' and doctors' case histories was made.

Results of evaluation would tend to diminish the criticisms

put forward of the 1969 system, that students are being taught quite effectively only to be more efficient and effective data gatherers. It would seem that they are also developing skill in evaluation of that data using the simulation technique which may imply development of skills of combining data. However, the simulations are still very low fidelity and whether or not students are developing a transferable or generalisable skill is unknown. It may be that the presentation of the case histories in a standardised format assists students' conceptualisation of a history and organisation of data, and that this is indirectly measured as skills in data evaluation. But we cannot say whether or not students are better problem solvers or more skilled in clinical reasoning. Murray et al (1977) report an improvement in factual knowledge retention and management decision making skills using computer assisted learning in Glasgow, but the report is not specific enough to enable any idea of how the teaching occurred or of what is meant by "management decision".

In any attempt to teach aspects of the diagnostic process, a model of this process is either formally or implicitly acknowledged by the teacher. Such a model may be either 'normative', setting out the ideal norm against which people try to make better decisions, or 'descriptive', attempting to represent the decision making behaviour of clinicians. (Taylor, 1972). This fact represents a considerable hazard in any computerised approach, since either represents a degree of inflexibility which may be inappropriate to the student's cognitive processes. In other words, the structures of teacher or teaching method and learner may well conflict, whether these be structures of substance or aptitude. Schneiderman and Muller (1972) report on their Diagnosis Game, which uses relatively simple remedial loops based on a normative version of a case provided by a clinician. The aim of this game is "the development of problem solving skills, free of the constraints of the real world". The loop involves the student typing in his revised differential diagnosis, ordering more data and finally making a definitive diagnosis. The computer gives an evaluation of

the diagnosis and, if necessary, explains why it is incorrect. This format obviously places fewer constraints on the students' thinking processes than does the ODSAL formulation, and so may allow greater transferability of the problem solving skill to real situations.

Harless et al (1971) report a similar normative simulation called CASE (computer aided simulation of the clinical encounter). In CASE sessions, the student types in requests, with no cuing or ordering constraints. When he has enough information, he enters the treatment phase and is requested to give his diagnosis and treatment plan. No pedagogical interruptions are made during the problem, and feedback is limited to a descriptive analysis of performance and comparison with optimal problem management. The analysis of deviations from the norm is much more flexible and detailed than in the Diagnosis Game. No evaluation is given of the specific problem solving skills being trained, nor of transferability of these skills to the real clinical situation.

Taylor (1972) reports a normative model based on Bayes' Theorem in which students are encouraged to estimate their prior probabilities at each stage of the diagnostic process. It is unknown to what extent this method enhances the cognitive problem solving skills applied in the clinical situation. Newble (1975) reports other computer applications including computerised patient management problems.

Evaluation of computer assisted training programmes in terms of the criteria of structure, transfer and problem solving is dependent upon the specific approach and components of the programme in question. The linear nature of the programmes and interruptions for use of the type facility, ensure that practice in clinical problem solving as it would occur in a clinical encounter, is not given. But this hazard applies to all the teaching methods so far described. Indeed, this would be an unreal aim, since the student in each case is undertaking

a deliberate learning task in the way in which a clinician is not.

Those programmes, such as Taylor's (1972), which require students to estimate their numerical prior probabilities are clearly working within a structure of aptitude which would not pertain in clinical practice and would only transfer, if at all, in a very mutated form. Such a programme does not take advantage of the student's normal thinking processes and is structurally inappropriate, unless its aim is to teach students to think probabilistically in numerical terms. But the type of more recent programme which de Dombal describes (1972; 1974) may well alert the student to some analysis of the clinical encounter and his own dynamic, problem solving, data gathering role in it. Such awareness may well transfer to the real clinical situation if bridging links of some type are provided. As far as direct training in problem solving thinking is concerned, again this method provides only an indirect, mediated content. The student's original structure of knowledge is not challenged or tested for its appropriateness, although structural reorganisation, or flexibility, may occur simply as a result of working through the simulation in a conscious, analytical manner. Thus, computer assisted training programmes seem, as yet, not to have addressed some of the basic and fundamental issues for those who wish to teach diagnostic thinking processes or clinical problem solving methods.

4.3 Summary and Conclusions

Three concepts have been identified and discussed which seem relevant and fundamental to teaching concerned with the development of the diagnostic thinking process. These concepts are: structure in learning; transfer of learning; and the relationship between problem solving and learning. It is suggested that these three, at a minimum, must be taken into account when devising teaching strategies for development of the diagnostic thinking process.

Strategies already devised and implemented are then discussed. These are: the integrated curriculum; problem based learning; training in hypothesis generation; cognitive skills training; and, computer assisted training. Each of these is evaluated in the light of the three fundamental concepts of learning. Each method is seen to have both advantages and disadvantages and deal with the development of the diagnostic thinking process with varying degrees of immediacy and directness.

On the basis of our discussion, description and review, we may draw a number of conclusions:

Firstly, knowledge must be presented, and assimilated by the student in a way that will encourage structuring appropriate to its future use in the clinical problem solving context and process.

Secondly, knowledge must be structured such that transfer of substance is facilitated across learning experiences and clinical problems.

Thirdly, the structure of knowledge must be sufficiently flexible to enable cognitive reorganisation as new information is learned.

Fourthly, given the range of individual and case differences, it is not appropriate to specify the ideal diagnostic thinking process in other than general, cognitive terms of aptitude, not substance. The diagnostic thinking process occurs in an unpredictable environment, therefore that process itself, in its substance and aptitude must be flexible. It is considered that teaching strategies must present that process as an entity amenable to conscious awareness and monitoring but not necessarily to particular shaping other than correction should cognitive error occur. Such presentation of principles, possibilities and ranges of thinking process should ensure transfer of learning to any circumstance in which such thinking occurs.

Fifthly, our knowledge of the diagnostic thinking process, even in its current general form, enables the aptitude to be treated as substance for the purposes of teaching and learning while, simultaneously, permitting development of the aptitude itself.

CHAPTER FIVE

Research Design: Hypotheses, Research Methods, Subjects, Design Validity

This chapter presents a statement of the research hypotheses of this study. A broad description of the two research methods is given, while details of the development and data analysis of these is reserved for the subsequent two chapters. The description of research methods is followed by details of the subjects of the study. The chapter closes with a discussion and evaluation of the validity of the research design, and a statement of the acceptable level of statistical significance for the study.

5.1 Hypotheses

Before presenting a statement of the research hypotheses, it may be useful briefly to consider their derivation. Elstein et al (1978), in discussing the diagnostic thinking process, state that:

"There are obvious dangers in allowing hypotheses and conjectures to influence data collection and interpretation at an early stage. These dangers include possibilities of premature closure, selective information gathering and biased interpretation of information". (p. 253)

The same could be said of research hypotheses, and so a large number of the present hypotheses, although framed in null terms, are asking questions rather than making predictions (see Hypotheses 5 to 9 and 17 to 20 inclusive). As far as content of hypotheses is concerned, psychological theory presents myriad potential explanations and models of behaviour, but no established criteria for selecting among them (Paxton, 1976) or for amalgamating different explanations (Pound, 1978). The hypotheses are, therefore, largely empirical rather than theoretically committed. We consider that "the macroscopic structure of the whole" (de Groot, 1965) is defined by 'hypothesis generation and testing' and that research hypotheses may be built upon this and upon the

criticisms made and questions asked of this model (see Chapter Three) and upon our expectations of its corollaries. We may now consider these hypotheses.

With regard to the questionnaires in endocrinology and neurology, which have the following sections:

- A. Mastery of factual knowledge.
 - B. Interpretation of symptoms and signs.
 - C. Selecting and testing diagnostic possibilities.
 - D. Formulating a diagnosis.
1. Accurate interpretation of symptoms and signs is the primary pre-requisite for diagnostic acumen, therefore scores on section B will have more predictive power than scores on sections A or C for scores on section D.
 2. Mastery of the separate skills tested in sections A, B, C and D is presumed to develop differentially during medical education and clinical practice. Therefore, statistically significant interaction effects will be found between groups of subjects and sections of the questionnaire.
 3. Given the presumed greater development of skill in sections B and D with clinical practice rather than medical education, observed differences between students' and registrars' scores on these sections will be greater than those observed on sections A and C. However, the greater experience and postgraduate learning of the registrars will cause groups' scores on all sections to be statistically significantly different in favour of registrars.
 4. Sections A, B, C and D test different skills which may be rehearsed in medical education and clinical practice respectively to disproportionate degrees. Therefore, for each group separately, statistically significant differences will be found between scores on the four sections.
 5. Scores on sections A, B and C combined have no predictive power for scores on section D, either in endocrinology or neurology.

6. No differences will be apparent between students and registrars in the relationship between scores across sections A, B, C and D, either in endocrinology or neurology.
7. No major differences will be observed between results for endocrinology and neurology.
8. No developmental pattern will be apparent, such that differences in scores between students and registrars will not be interpretable in terms of differences between medical education and clinical practice.
9. Results will have no interpretative value for current descriptions of the diagnostic thinking process

Note Hypotheses 5 to 9 are framed in null terms as a matter of formality only. They could equally as well have been couched as questions.

With regard to the accounts by stimulated recall:

10. (a) Students, house officers and registrars will make multiple pre-diagnostic and diagnostic interpretations of clinical information as it is progressively elicited during the clinical interview.

(b) No differences will be observed between students, house officers and registrars in the relative use of pre-diagnostic and diagnostic interpretations of clinical information during the entire diagnostic thinking process.
11. (Elstein et al (1978) find that, in terms of time of onset, the first diagnostic hypothesis is generated by ten per cent of the way through or within the first five minutes of the clinical interview. Barrows et al (1978) find that the first diagnostic hypothesis is advanced, on average, within 28 seconds of the appearance of the main complaint).

The present study will show that students, house officers and registrars make immediate interpretative or evaluative response to initial items of clinical information received whether or not these actually constitute the patient's

main complaint.

12. Accounts given by students, house officers and registrars will indicate that the diagnostic thinking process involves working within a cognitive context or contexts extrapolated from the clinical information available.
13. It will be observed for students, house officers and registrars that reinterpretation of clinical information occurs during the course of the clinical interview due to:
 - (a) New thoughts occurring about already interpreted clinical information when no new information has been added.
 - (b) New clinical information being elicited to facilitate reinterpretation of clinical information already elicited and interpreted.
14. No differences will be observed between students, house officers and registrars in use of strategies for selection between competing interpretations of clinical information.
15. Given that all subjects commence their clinical enquiry with a question intended to elicit the patient's presenting complaint, the course of the clinical interview can be determined by the following factors:
 - (a) The flow of information as presented by the patient.
 - (b) The flow of information as elicited by the subject according to his interpretations of the clinical information.
 - (c) The logical structure of the standard (taught) clinical history.

No differences will be observed between students, house officers and registrars in their use of (a), (b) and (c) in determining the course of the clinical interview.

16. On review of the clinical interview, students, house officers and registrars will identify areas of omission in the information elicited from the patient. Such

omission may be in two areas:

(a) Specific enquiry directed at the patient's problem, symptoms and signs or arising from the subject's interpretation of the clinical information elicited.

(b) General or routine enquiry.

17. Results will not indicate the mechanism of interpretation and reinterpretation of the array of clinical information as it is accumulated throughout the clinical interview.
18. Results will not indicate mechanisms of actual or potential error in the diagnostic thinking process.
19. Results will not indicate any categories or types of information other than those of the standard (taught) clinical history, used to assist either in pre-diagnostic or diagnostic interpretation of clinical information or in selection of the most likely diagnosis.
20. Results will provide no indication of the nature of psychological probability in the diagnostic thinking process.

Note Hypotheses 17 to 20 are framed in null terms as a matter of formality only. They could equally as well have been couched as questions.

5.2 Development of the Current Approach

This second section will present a brief introductory discussion followed by a description of an initial and unsuccessful attempt to draw up clinical problems to test formal thinking processes. The section concludes with the rationale for the research approach finally adopted. The subsequent two sections (5.3 and 5.4) describe the research methods and rationales.

Note During the stages of design, development, implementation and analysis of research methods and results, Dr. Leslie Sedal, Clinical Tutor and Consultant Neurologist, and Dr. Philip Marsden, Clinical Tutor and Consultant

Physician specialising in endocrinology, advised on the technical medical content of the work and used their good offices to procure many of the subjects of the study. Dr. Sedal advised on the questionnaire in neurology. Dr. Marsden advised on the questionnaire in endocrinology, the clinical problems in propositional logic and the account gathering study.

5.2.1 Introduction

Dudley (1971) points out that studies of the clinician in action had, until that time, been either stylised or were only just beginning. He suggests a number of reasons for this:

"... first, the matter is a complex one and has awaited hypotheses that can direct inquiry; second, there is the distinct possibility that, in the context of a clinical encounter, clinicians feel threatened by an eavesdropper, particularly when they may feel guilty because their own efficient methods of pursuing a diagnostic pathway bear such a small resemblance to those they were taught in medical school; third, the experiment is likely to be an untidy one in that the intrusion of the observer cannot help but significantly influence both doctor and patients".

Despite, or perhaps because of, the conceptualisation of the process of 'hypothesis generation and testing' as the method of clinical problem solving, the first of Dudley's three factors appears substantially to have remained. Although objectives may be stated (Elstein et al, 1978), research hypotheses and questions rarely are defined, and research seems to have stagnated at the stage of quantification of aspects of hypothesis generation and testing, rather than advancing towards a deeper understanding, either practical or theoretical of that broad process. As for Dudley's second and third factors, a description of the present research provides an account of attempts to overcome these problems. Firstly occurred an unsuccessful attempt.

5.2.2 Clinical Problems in Propositional Logic

Our initial approach to the study of clinicians' diagnostic

thinking processes was based on research hypotheses stemming from the obvious theoretical contiguity between the identified process of hypothesis generation and testing and the Piagetian stage of formal operations. Our intention was to test and define the relationship between formal operational thinking in everyday situations and the same in the solution of clinical problems, our hypothesis being a developmentally related one of a relationship between logical thinking in everyday life and in clinical problem solving. Our main frame of reference, having taken this approach, was a special application and circumstance of Piagetian theory rather than clinical problem solving thinking per se. We therefore commenced by identifying the sixteen binary operations which characterise formal operational thinking. These operations are as follows (Inhelder and Piaget, 1958):

1. Conjunction ($p \cdot q$) p is true and q is true simultaneously.
2. Non-implication ($p \cdot \bar{q}$) p is true and q is false.
3. Negation of q ($\bar{q} [p]$) q is false whether or not p is true.
4. Affirmation of p ($p [q]$) p is true whether or not q is true simultaneously.
5. Disjunction ($p \vee q$) Either p is true or q is true or both are true.
6. Negation of p ($\bar{p} [q]$) p is false whether or not q is true.
7. Affirmation of q ($q [p]$) q is true whether or not p is true simultaneously.
8. Incompatibility (\bar{p}/\bar{q}) Either p is false or q is false or both are false.
9. Conjunctive Negation ($\bar{p} \cdot \bar{q}$) Both p and q are false.
10. Exclusion ($p \vee \vee q$) Either p is true or q is true.
11. Inverse Non-implication ($\bar{p} \cdot q$) q is true, p is false.
12. Equivalence ($p = q$) Either p and q are true simultaneously, or neither is true.

13. Implication ($p \supset q$) If p is true, q is true; but if p is false, q may be true or false.
14. Inverse Implication ($p \subset q$) If q is true, p is true; but if q is false, p may be true or false.
15. Complete Affirmation ($p * q$) Admits of any binary combination of p , \bar{p} , q and \bar{q} .
16. Complete Negation (0) Admits of no binary combination of p , \bar{p} , q or \bar{q} .

These sixteen combinations result from taking as base elements the four associations and combining them one by one, two by two, and so on, forming a set of all possible combinations or a lattice. (For a full discussion see Flavell, 1963).

On the basis of the sixteen binary operations, construction of clinical problems was attempted, in the form of:

- a) Statement.
- b) Subject's action and reason.
- c) Result.
- d) Repeat (b) and (c) to closure.

The problems were to be presented to the subject in a manner analogous to de Dombal et al's (1972b) VERBAL mode. Each problem was developed by taking a clinical example of the binary operation in question and working backwards. For example, $q [p]$ may be demonstrated in that a diagnosis of diabetes can be made whether the fasting blood sugar level is normal or raised as long as the glucose tolerance test (G.T.T.) is abnormal. On the basis of this, a clinical problem of the following type was developed:

"A patient is found to have glycosuria and the blood glucose is estimated but found to be normal. The patient is reassured and goes back to work.

Has this patient had adequate investigation and treatment?

Two weeks later the doctor again checks the fasting blood glucose and finds it elevated at 300mg %. He refers the patient for a G.T.T.

Do you regard this as sensible? Give reasons.

The G.T.T. result is normal. The patient is now firmly convinced that he is a diabetic.

Are you also convinced? Give reasons".

For an example of a problem at a different stage of development, let us take $p.\bar{q}$, the truth of which implies the falsity of $\bar{p}.\bar{q}$, $p.q$ and $\bar{p}.q$ (and vice versa). In developing this problem, p was taken to be a grand mal seizure (or evidence compatible with or suggestive of this) and q was taken to be co-ordinated (cortical) cerebral activities such as communicating speech or co-ordinated motor activity or memory. Prior to solution of the problem, the subject is given all necessary information about the features of syncope and epilepsy (see Appendix 1). The operation in question, then, may be established by the subject testing the following propositions:

To be proved true:

$p.\bar{q}$ - evidence of grand mal seizure, no speech possible.

To be proved false:

$p.q$ - evidence of grand mal seizure, with communicating speech.

$\bar{p}.q$ - no evidence of grand mal seizure, with communicating speech.

$\bar{p}.\bar{q}$ - no evidence of grand mal seizure, but not able to speak.

Some problems did not take the form already described, but were presented more simply if appropriate. For example, when considering \bar{p}/\bar{q} , a problem was presented for solution as follows:

"A patient says that she has not passed water for 5 days but has continued to drink normally. A weight chart recorded by new nurses shows that her daily weight has apparently not changed. What is your action?"

This impossible finding represents $p.q$ which is therefore false. Any of three other combinations may be true and each should be considered by the subject, as follows:

$\bar{p}.q$ - patient incorrect, chart correct.

$\bar{p}.\bar{q}$ - patient incorrect, chart incorrect.

$p.\bar{q}$ - patient correct, chart incorrect.

Testing, or having the intention of testing, each of the above combinations would evidence formal operational thinking in the operation of compatibility only.

Such clinical analogues were defined for each of the sixteen binary operations. However, the proposed format of the problems was found to be unserviceable in that it was not proved possible to construct a problem to test for the presence of a certain thinking operation without having that problem encourage the subject to think in the manner being tested. In other words, due to the structure of the content, the problems had the apparent propensity to guide the problem solver into certain ways of thinking, regardless of whether or not he would have done so when unguided. We concluded that this intractable problem resulted from an unresolved conflict about the aims and paradigms of the research, in essence about the relative importance of the theory of formal operations and the process of clinical problem solving.

5.2.3 Development of the Current Approach

The resolution of the conflict described entailed giving primacy to the process of clinical problem solving and using psychological theory to interpret our findings. Such a resolution, then, implied a deeper study of the diagnostic thinking process itself, rather than testing research hypotheses about certain selected theoretical psychological approaches. This approach implied recognition of the possible need to embrace psychological and educational theory more generally for its varied explanatory, descriptive and theoretical interpre-

tative frameworks. It was considered that Piagetian theory may only account for part of the process; for example, the framework of formal operations may account for the nature of hypothesis generation less well than it accounts for the psychology of hypothesis testing. We therefore recommenced with the point of view that other psychological and educational theories may be useful, and that the former, in particular, would most usefully be identified and defined during the course of the research rather than prior to the stage of data collection.

Having adopted this approach, the research hypotheses accordingly changed in emphasis and became more closely based on the conclusions of the literature review. In particular, the hypotheses are intended to elicit answers to some of the questions put in 3.1.1 and 3.1.2 above. In parallel, it was decided to make a study of the development of diagnostic thinking processes from the final year of undergraduate medical education until the years of clinical practice at registrar level. Our research design, therefore, is based on three groups of subjects: final year clinical medical students, house officers and registrars. The questionnaire study, however, omits house officers, and the groups of subjects are different for the questionnaires and the account gathering study. This problem is discussed below (section 5.6). Comparison of these groups thus gives an indication of developmental aspects of the process under study and of the goodness of fit or appropriate relationship between undergraduate medical education and clinical practice.

5.3 Questionnaires in Endocrinology and Neurology - Rationale and Description

This section discusses the intended research purpose of the questionnaires in endocrinology and neurology which form one half of the present study. In addition, a brief description of the format and structure of the questionnaires is given.

Thus the questionnaires are here discussed from the point of view of the overall research design only. Details of specific content selection, format and validation are given separately below (see Chapter Six).

5.3.1 Rationale

Each questionnaire comprises four sections as follows:

Section A: Mastery of factual knowledge.

Section B: Interpretation of symptoms and signs.

Section C: Selecting and testing diagnostic possibilities.

Section D: Formulating a diagnosis.

It is taken that the sections of each questionnaire sample all the categories of teaching and learning extant in undergraduate medical education, with the exception of the physical examination.

The research aim of each questionnaire is to supply quantitative information concerning the relative roles and contributions of the skills tested in sections A, B and C in relation to diagnostic ability tested in section D, for students and practising, experienced clinicians separately. In addition, the results of these two groups of subjects may be compared in order to gain an indication of the effects of medical education and clinical practice on the relative roles and contributions of the skills tested. Having parallel questionnaires in two specialities enables a comparison of all results across different content, in order to test the content specificity or generalisability of findings, this latter point having far reaching implications for the structure of medical education.

The specialities selected for the comparative study of the effects of differing task environments were endocrinology and neurology, both of which are sub-specialities of internal or general medicine, but only one of which (neurology) is commonly taught as a speciality in its own right within the medical

school curriculum (General Medical Council, 1977). Both specialities are clinically based, although endocrinology may rely to a greater extent on laboratory investigations of chemical pathology. However, the greatest difference between these two specialities, apart from their concern with different bodily systems, may, perhaps, be seen as methodological and pedagogical. Although no evidence can be cited, it is probably the case that neurology has the reputation of being one of the most difficult specialities of internal medicine, and yet in no other speciality is the process of diagnosis made more explicit or a logical approach more emphasised. The G.M.C. survey (1977) makes this clear:

"It is hoped that the neurology course will reinforce students' basic medical science knowledge and reinforce their appreciation that this knowledge must be applied consciously and systematically in diagnosis and investigations ... Great stress is laid on the need for a full history and a full description of all findings".

No other sub-speciality of general medicine states such a clear philosophy. In neurology, students are traditionally taught to:

- a) localise the lesion, i.e. from the symptoms and signs, decide which precise part of the nervous system has been damaged;
- b) define the general pathology, i.e. define the broad class of disease - infection, cancer, etc.;
- c) define the special pathology, i.e. identify the exact infectious organism involved, or the nature of the cancer, etc.

Some key tests in neurology require general anaesthesia and occasionally have serious complications, therefore students are normally taught to consider carefully which investigations will be helpful and to come to as accurate a diagnosis as possible on the clinical evidence before ordering tests.

The nature of endocrinology does not, routinely, lend itself

to displaying any of these characteristics, being less structured and formalised in its diagnostic surface structure and having clinical manifestations of a more protean type. The coincidence of similarities and differences makes neurology and endocrinology appropriate specialities for comparison, being neither too dissimilar and therefore prone to yielding unrepresentatively large and possibly ungeneralisable observed differences, nor too similar and so not evidencing real differences in test results. In addition, the traditional teaching of, at least, the surface structure of neurological diagnosis, makes this difference a potentially interesting one for this study.

5.3.2 Description of Format and Structure

Both questionnaires follow the same pattern of structure and content. Each consists of four sections, three of which are made up of five-option independent true-false multiple choice questions. The remaining section requires statement of the most likely diagnosis for a given case history. The skills tested in sections A, B and C as described above (5.3.1) are possible components of the skill tested on section D. See Figure 8.1 below for a diagrammatic representation of the relationship between sections of the questionnaire. Appendices 2 and 3 give the precise content of each questionnaire.

5.4 Accounts by Videotape Stimulated Recall

The study of account gathering by videotape stimulated recall of the diagnostic thinking process during a clinical interview is separate from the questionnaire study just described, but complementary to it, measuring or tapping different aspects of the process under study.

5.4.1 Rationale

The questionnaire study described above yields a quantitative indication of the power or weight of the various skills tested in the first three sections in predicting the scores of subjects

on the test of making a diagnosis. The study also yields information concerning subjects' relative mastery of the four skills tested. However, such data do not and cannot indicate the manner of cognitive functioning. It is very much of the 'black box' type, such that the experimenter is aware of input and output, but not of the nature of the mediating, causal or dynamic process which connects the two. A further weakness of the questionnaire as a research instrument in this field is its low-fidelity nature. It cannot be argued that responding to a written case history is really like interviewing a patient and forming a diagnostic opinion. The dynamic, contextual aspects of the clinical interview are entirely missing. The subject has no control over data presented in a questionnaire, neither is he subject to the many sources of cues and distractions, difficulties and facilitatory aspects of the real clinical interview. And if he were, a questionnaire could not tap his responses or the dynamic of his thinking. In brief, the diagnostic thinking process, as such, in its richness, with all its vagaries will not be reflected in the results of an objective questionnaire. Therefore, having achieved measurement of different skills and their predictive values, some other research instrument is required to indicate the cognitive processes which yield those predictive values.

Given that the diagnostic thinking process occurs only under conditions where a diagnosis must be made, then a high-fidelity research instrument was required. The experiences of the Michigan and McMaster groups described above (sections 3.1.1 and 3.1.2) indicate the usefulness of the method of stimulated recall and provide a vicarious research experience, demonstrating the strengths and weaknesses of the method. It is considered that neither group has yet produced an adequate, reasoned assessment of the method as a research instrument and each has, therefore, introduced sources of invalidity or unreliability as described above also (sections 3.1.1 and 3.1.2). Given its apparent high-fidelity nature and the possibility of providing an established validation and reasoned assessment

of the method, and thus armed with the experience of others, a form of stimulated recall was selected as appropriate to the demands of this study. Chapter Nine below presents the full discussion and establishment of method validity.

5.4.2 Description

The subject (either student, house officer or registrar) interviewed a patient in order to establish a diagnosis as far as possible and reasonable. The clinical interview was videorecorded. Immediately following the completion of the interview the subject and the experimenter reviewed the videotape in short sections while the subject commented in detail upon his own thoughts as they had occurred while interviewing the patient. The subject-experimenter phase was audio-recorded and a transcript prepared of the dialogue and content of each short videotape section replayed. The transcript was then subjected to content analysis. Chapter Nine below presents full details of the method.

5.5 Relationship Between the Questionnaire Study and the Account Gathering Study

It must be emphasised that the two parts of the present research are separate, but complementary, and have separate samples of subjects. It is argued (section 5.6) that the groups of subjects may be treated as samples from the same population where they are of the same category and as successive samples taken at intervals of time from the same population where they are of different categories (see 5.6 below). This accepted, the two research instruments may be seen as focussing on different aspects of the same process.

The questionnaires provide a quantitative baseline, against which the qualitative results of account gathering may better be interpreted. But the questionnaires provide only a static measure of a dynamic process. The questionnaires approach the diagnostic thinking process from a more analytical point of view than the more global approach of the accounting gathering.

The types of variable which each research method addresses are also different in that account gathering cannot measure knowledge, but the greater difference is found in the manner in which each method approaches its variables. The questionnaires measure absolute levels and highlight relationships between variables in a manner which fails to reflect the manner of use of variables and their changing or dynamic nature. These latter aspects are the domain of account gathering.

Thus the complementary research methods of this study provide information about the content and structure of thought, about its substance and its process. Neither method alone could achieve this wide spectrum. The two together, however, should facilitate a more full and rounded account of the diagnostic thinking process than has hitherto been reported. Pool (1959) summarises this point:

"It should not be assumed that qualitative methods are insightful, and quantitative ones merely mechanical methods for checking hypotheses. The relationship is a circular one; each provides new insights on which the other can feed". (p.192)

5.6 Subjects

It has been stated that the subjects of the two studies form separate samples which can be treated as though coming from the same populations. This section describes separately the subjects for each study and their comparability. It closes with a discussion of the comparability of the subjects of the two separate studies.

5.6.1 The Questionnaire Study

Table 5.1 shows details of age and sex for each group of subjects for each questionnaire.

The subjects for each questionnaire are 35 final year clinical medical students and 35 post-membership (Royal College of Physicians) registrars. The students were from four London

Table 5.1 Questionnaires in Endocrinology and Neurology:
Age and Sex of Subjects

Group	Number in Group	Sex			Age Range in Years	Mean Age	Standard Deviation
		M	F	?*			
Endocrinology students	35	21	14	-	22 - 31	22.97	1.69
Endocrinology registrars	35	33	1	1	28 - 46	30.15	3.93
Neurology students	35	16	11	7	22 - 39	23.5	3.35
Neurology registrars	35	30	1	4	28 - 42	30.44	3.21

* Refers to subjects who failed to give personal details of age and sex.

and two provincial medical schools and the registrars from an additional London hospital and many provincial centres. Our results, then, are based on a wide range of backgrounds and experiences; within group variation is therefore taken into account in the validation procedures applied to the questionnaires. However, the heterogeneity of backgrounds and experiences is considered a positive feature of the research design, likely to randomize any effects of special environments and to reveal common features across all subjects.

Since the two groups of subjects are to be treated as though they are a longitudinal sample when, in fact, they are two separate cross-sectional samples, we must consider the legitimacy of this approach by proving the comparability of the samples. In essence, it is necessary to show that if any differences are found between students and registrars these are due to development of the latter group from a state similar to that found in the former group. This is a reasonable contention.

Although medical education is constantly changing, developing and improving, the changes which have occurred are not such as would alter the processes which concern the present study. Most changes in medical education have occurred in the fields of medical teacher training, attitude measurement, teaching methods and assessment, as a brief review of the journal of the Association for the Study of Medical Education reveals. Changes in curriculum content or characteristics of entering medical students are rarely reported. Isolated changes may occur in individual medical schools, but it cannot be said that any sweeping and general changes in curriculum content or student characteristics have occurred during the past decade. The move towards integration discussed in Chapter Four is probably the most widespread of changes in that most medical schools now attempt some form of integration, often on a very small scale. It is argued, however, in Chapter Four that while transfer of learning may be facilitated by the integrated curriculum and thereby produce a positive, but mediated, effect upon the diagnostic thinking process, that process itself is unlikely to be altered and overall curriculum content does not necessarily change. We would argue, then, that it is a reasonable analytical approach to take our students and registrars as though they are a longitudinal sample, and attribute differences to the relative effects of medical education and clinical practice, since no substantial changes have occurred in the former since the registrars were themselves students. In addition, all groups of questionnaire subjects are from both London and provincial areas. We may conclude that comparison of student sample with registrar sample is quite in order and that differences may be attributed to the relative experience of medical education and clinical practice. In addition, we may conclude that comparison of subjects in endocrinology with those in neurology is also permissible since no identifiable group differences exist between the student samples respectively.

5.6.2 The Account Gathering Study

The subjects who gave accounts by stimulated recall were:

22 final year clinical medical students

22 pre-registration house officers

22 post-membership (Royal College of Physicians) registrars

The students were taken from one London medical school, while the house officers and registrars were from a London teaching hospital and a London district general hospital. Distribution by sex within each group is as follows:

Students: 7 female, 15 male

House Officers: 6 female, 16 male

Registrars: 2 female, 20 male

Details of age are available for students only (in years):

Mean age: 23

S.D.: 1.8

Range: 22 - 30

In terms of years of clinical practice, house officers, by definition, have less than one year of experience. The group of registrars' years of clinical practice are as follows:

Mean: 5.2

S.D.: 1.4

Range: 3 - 7

As with the questionnaire study, the groups of subjects are to be treated as though they were a longitudinal sample when, in fact, they are separate cross-sectional samples. The arguments discussed in 5.6.1 above, therefore, also apply to the account gathering study. However, an additional point must be made in order to show the comparability of the group of house officers and the group of registrars. This is to emphasise that not only have no substantial changes occurred in medical education since the clinicians were students, but also that no substantial changes have occurred in clinical practice since the group of registrars qualified. In other words, we must point to the fact that the clinical experience of the house officers matches that

of the registrars when they were house officers themselves. This being so, then we may again conclude that our treatment of our three groups of subjects as a longitudinal sample is quite justified.

5.6.3 Comparability of Questionnaire and Account Gathering Subjects

Since the two research methods are to be used in a complementary manner, each to add interpretative value to the other, then comparability of subjects across research methods must be established. Groups of subjects are as shown in Table 5.2.

<u>Table 5.2 Subjects of the Complementary Studies</u>			
Subjects	Questionnaires		Account Gathering
	Endocrinology	Neurology	
Students	35	35	22
House officers	-	-	22
Registrars	35	35	22

From the point of view of the research design, there are two parallel research methods but three parallel research samples, in that the subjects for the questionnaires in endocrinology and neurology were also separate samples. We have, however, established the longitudinal comparability of samples within methods. Considering the comparability of samples across studies, all students were in their final clinical year, while all registrars had gained their membership of the Royal College of Physicians. The only discernible, but minor, difference is that questionnaire subjects are from both London and provincial areas, while all account giving subjects are from London only. However, this difference is lessened in that not all registrars and house officers had initially trained in London. It is not considered that this constitutes any jeopardising degree of variability with regard to the subject matter of this study.

Neither medical education nor clinical practice varies across the country to any invalidating extent. It is therefore argued that conservative comparison of results of the complementary studies is within the limits of the research design.

5.7 Design Validity

A research design is successful to the extent that it accounts for such extraneous variables as may bear upon the research, and to the extent that it allows for "optimal statistical efficiency" (Campbell and Stanley, 1963). All research is concerned with the relations between variables and in developing the design for research Ferguson (1966) suggests that:

"... the experimenter must (1) select the values of the independent variable, or variables, to be compared; (2) select the subjects for the experiment; (3) apply rules or procedures whereby subjects are assigned to the particular values of the independent variable; (4) specify the observations or measurements to be made on each subject".

In addition, in order to calculate the degree to which results are generalisable, the research must be designed in such a way that it lends itself to appropriate data analysis. The precise method and type of analysis must be carefully chosen - in effect the methods to be applied in the analysis of the data and the type of questions which research can answer are determined by the nature of the variables.

This short discussion, then, has introduced three major areas to be considered in any research design - internal validity, external validity and data analysis. This section deals with the first two of these. Data analysis is discussed in the two chapters which the development of the questionnaires (Chapter Six) and of the account gathering method (Chapter Nine).

5.7.1 Internal Validity

A research design is internally valid if the factor to which results are attributed is the only one possible, and if those

results themselves are reliable and valid. That is to say that all extraneous variables which may influence the result or cause results to be ambiguous must be controlled or accounted for. Reliability and validity of data are discussed and established below (see Chapters Six and Nine). We may determine the extent to which the present design is internally valid by considering the factors which may jeopardise it. Campbell and Stanley (1963) cite eight such factors.

The effects of history, maturation and testing are concerned with designs of the test-treatment-retest type and are not of relevance to the present design in which each group of subjects is studied only once, and comparisons made between groups. Likewise, the effects of statistical regression are not relevant to the present design. Instrumentation, however, is a relevant variable. This jeopardising effect occurs when changes in the calibration of a measuring instrument or changes in the observers or scorers used produce changes in the obtained measurement. Multiple choice questions are not subject to such a factor, scoring being purely objective. Our discussion of the method of account gathering by stimulated recall (Chapter Nine), however, indicates that instrumentation is seen as one of its major potentially jeopardising extraneous variables, in the sense that slight differences in the implementation of the method may cause large differences in the depth and quality of accounts. For this reason careful attention was paid to standardisation of the method and the same experimenter gathered all 66 accounts. Jeopardising effects of instrumentation have therefore been minimised.

A second possibly jeopardising factor is found in biases concerned with differential selection of respondents for the comparison groups. Campbell and Stanley (1963) discuss this variable in relation to groups which are taken as comparable. The present design, for each of its separate research instruments, deliberately compares groups of subjects contrasting in their amounts of experience of medical education and clinical practice, but similar where those experiences overlap as shown in 5.6.1

and 5.6.2. However, the jeopardy of the selection variable must be considered in relation to the admissible degree of comparison between the parallel, complementary studies. This is discussed in 5.6.3 above and comparability of samples established. The selection variable, therefore, is not a jeopardising factor.

Campbell and Stanley's (1963) final two possibly jeopardising variables, experimental mortality and interactions between the variables, do not apply to the present research design. It is, therefore, considered that internal validity is demonstrated.

5.7.2 External Validity

Research has external validity to the extent that its results are generalisable to other sections of the same population and also to other populations. If the results are specific only to the sample studied, for whatever reason, then the design has no external validity. We may determine the extent to which the present design is externally valid by considering the factors which may jeopardise it.

There are two possible factors which may jeopardise the external validity of the present design; these are the representativeness of the samples and the reactive effects of the research methods.

There would seem to be little reason to suppose that the present samples are unrepresentative of British final year medical students, house officers and medical registrars, although we may not be justified in generalising to registrars in surgical specialities.

There are differences in the structure of medical education from school to school in Britain, for example in the degree of integration of clinical and pre-clinical curricula or of separate specialities, and some of our subjects from provincial schools and hospitals have come from schools which have been particularly pioneering and innovative in this way. However,

our validation processes for the questionnaires show a degree of within group homogeneity which may indicate that the products of these differing structures themselves differ, if at all, in areas other than those tested (see Chapter Six). We may therefore presume adequate representativeness of our present samples of the population of final year medical students, house officers and medical registrars.

Considering the second possible jeopardising variable, any reactive effects of the research methods themselves would preclude generalisation of findings to populations not exposed to those methods. In other words, there may be the possibility that results are artefacts of the research method. The developmental and validation procedures applied to the questionnaires, and wide use of such methods for assessment in both undergraduate and postgraduate medical education would suggest appropriateness of the task and familiarity of the method to the subjects. These are the only possible precautions to be taken against reactive effects. It is also shown that the accounts are based on a considerable degree of realism in a familiar situation, while Chapter Nine also shows that most subjects felt that they had experienced no reactive effects of the research method.

To this extent, then, we may be confident in the validity of generalising our findings to the population from which our subjects are drawn. No conclusions may be stated about generalisability to other populations, but comparison of findings with those, for example, of the American and Canadian studies may give an indication of the feasibility of generalisation.

5.8 The Acceptable Level of Significance: $p < .01$

In deciding the acceptable significance level for any research study, it is necessary to decide how infrequently a result should occur before the null hypothesis may be rejected (Jolly and Gale, 1976). This decision may be based on a number of factors. In the present study it is considered likely, given

the differing levels of experience of the three types of subjects, that differences will occur in the population relatively frequently. In addition, the results of these research instruments themselves must be subject to conservative interpretation, on the one hand because of the low fidelity nature of one instrument (which is, nonetheless, both valid and reliable), and, on the other hand, because of the uncertainties of completion of data collection by means of the second instrument. Neither of these features is invalidating providing that data is interpreted conservatively. For these reasons, the acceptable level of significance for the results of this study is set at $p < .01$. This more rigorous level than $p < .05$ implies the possibility of relying with greater confidence on our observed results as real, rather than as caprices of sampling, research design or instrumentation.

Exceptions to this significance level occur with regard to development of the questionnaires. The reason for exceptions will be explained in the text. However, for all test results, the acceptable level of significance remains at $p < .01$.

5.9 Summary and Conclusions

A statement of the research hypotheses of the present study is given. The historical development of the current approach is described, including initial unsuccessful approaches. A broad description and discussion of rationale is then given for each of the two parallel research methods of the study which are structured questionnaires in endocrinology and neurology and account gathering by videotape stimulated recall. The relationship between these two methods is discussed.

The subjects of the studies are described and the rationale for horizontal and vertical comparability of samples presented. Finally, both the internal and external validity of the research design and the acceptable level of statistical significance are established.

CHAPTER SIX

The Questionnaire Study 1: Validity, Reliability and Data Analysis

This chapter presents a discussion of the identification, development and rationale of the content and structure of the questionnaires in endocrinology and neurology in terms of their associated validity and reliability.

In addition, the statistical methods of data analysis are described. It will be remembered from section 5.3 above that the questionnaires in endocrinology and neurology have the same components, each comprising four sections. Each section deals with a different aspect of the diagnostic process as follows:

- A. Mastery of factual knowledge.
- B. Interpretation of symptoms and signs.
- C. Selecting and testing diagnostic possibilities.
- D. Formulating a diagnosis.

Sections A, B and C comprise five option independent true-false multiple choice questions. For a discussion of the rationale for the selection of this format, see Appendix 4. Section D requires the subject to write the most likely diagnosis for each of five given case histories. Questions in section C are based on the case histories also.

There is some considerable repetition of topics across sections to enable interpretation of the relationship of scores across sections, each section testing a different aspect of clinical knowledge and (cognitive) skill. The content of the questionnaires is reproduced in Appendices 2 and 3. It will be noted that sections C and D are not separate within the questionnaires, although they are separated for purposes of analysis. Tables 6.1 and 6.2 summarise Appendices 2 and 3, showing the topics of each section, the total number of questions per section and the possible range of raw scores per section, for the questionnaires in endocrinology and neurology respectively. Where a diagnosis is made up of separate components, each part is awarded a separate mark. For a description and discussion of the instructions given to respondents, policy on guessing and the scoring schedule, see Appendix 5.

Section	Topics	Number of questions	Possible range of raw scores
A	Juvenile onset diabetes mellitus Hyperthyroidism Maturity onset diabetes mellitus Acromegaly Hypothyroidism Primary hyperparathyroidism Klinefelter's syndrome Cushing's syndrome Autoimmune thyroiditis Addison's disease Turner's syndrome Pheochrom cytoma	12	-60 to +60
B	Hypoglycaemia Hypercalcaemia Diabetes insipidus Infertility Short stature Gynaecomastia Delayed puberty Goitre Hypopituitarism Tetany Virilism Obesity	12	-60 to +60
C and D	Myxoedema Diabetes mellitus Klinefelter's syndrome Bronchial carcinoma with ectopic ADH secretion Primary hyperparathyroidism	C - 15 D - 5	-75 to +75 0 to +6

Table 6.1 Questionnaire in Endocrinology - Topics, Number of Questions, Range of Raw Scores Possible

Having described the content and structure of the questionnaires, we may now proceed to consider questions of validity and reliability. Although multiple choice questionnaires are located in the ranks of testing methods under the broad classification of objective tests, the actual degree of objectivity associated with any test is largely dependent upon the thoroughness of its design and development procedures, the appropriateness of its content and instructions to the respondent and item selection. The method of development and design

Section	Topics	Number of questions	Possible range of raw scores
A	Vitamin B ₁₂ deficiency Parkinsonism Migraine Multiple sclerosis Intracranial haemorrhage Cerebral tumour Epilepsy Myaesthesia gravis Meningitis Skeletal compression syndromes Alcoholism Drug therapy	12	-60 to +60
B	Upper and lower motor neuron lesions Cerebellar lesions Extrapyramidal disorders Sensory abnormalities Cranial nerve lesions Spinal nerves and roots Cerebral lesions Coma Raised intracranial pressure Neurological history	10	-50 to +50
C and D	Metastatic carcinoma in left cerebral hemisphere from primary focus in lung Haemorrhage from aneurysm of right communicating artery at base of brain Multiple sclerosis, lower thoracic spinal cord, T ₁₀ or slightly higher. Motor neuron disease, lower cervical, upper thoracic spine, T ₁ Tumour of 8th nerve, acoustic neuroma, right cerebello-pontine angle	C15 D 5	-75 to +75 0 to +16

Table 6.2 Questionnaire in Neurology - Topics, Number of Questions, Range of Raw Scores Possible

of the questionnaires took into account each of the above factors. Both questionnaires were subjected to the same procedure for establishing validity and reliability.

6.1 Validity of the Questionnaires

In constructing any test, the question of its degree of validity is of great importance. As Gronlund (1971) describes it:

"Validity refers to the extent to which the results of an evaluation procedure serve the particular uses for which they are intended". (p.75)

He qualifies this statement, pointing out that validity refers to results rather than to the instrument, that its presence is a matter of degree, and that it is always specific to some particular use. Guilford (1954) makes a further point that validity depends upon reliability. The establishment of reliability of the questionnaires is discussed in 6.2 below. Validity refers to the extent to which a test actually measures that which it purports to measure, while reliability refers to the consistency or repeatability of findings using the test. Validity may take a number of forms, depending upon the purpose of the instrument. Cronbach (1960) identifies the three major types, and the question they each answer, thus:

- (a) Criterion oriented validity, answering the question: How do measures of some valued performance (criterion) relate to test scores ?
- (b) Content validity, answering the question: Do the observations truly sample the universe of tasks or the situations they are claimed to represent ?
- (c) Construct validity, answering the questions: How can scores on the test be explained psychologically ? Does the test measure the attribute it is said to measure ?

Let us consider each of these in turn in relation to our present tests.

6.1.1 Criterion Oriented Validity

This may be of two types: predictive validity, when the second measure may be obtained at a future date; and, concurrent validity, when the second measure should be obtained on some present performance. For our purposes, a measure of concurrent validity was too difficult to attain. Gronlund (1971) warns

that problems "in locating a suitable criterion of success for the purpose of test validation are not unusual. The selection of a satisfactory criterion is one of the most difficult problems in validating a test." (p. 89)

Our tests contain four elements of medical skill - mastery of factual knowledge, interpretation of symptoms and signs, selecting and testing diagnostic possibilities, and forming a diagnosis. Our purpose is to discover the contributions of the first three of these to skill in the fourth. It follows, that the relationship between these elements has yet to be established. To use any one of them as the criterion is, therefore, inappropriate. The only remaining element of clinical skill is in performing the clinical examination of the patient. Again, the relationship of this skill to the four aspects already described is unknown, and valid, reliable measures of skill in examination of the patient are either difficult to achieve (Marshall and Ludbrook, 1972) or inappropriate to our situation, requiring resources and manpower not available to us (for example, Harden et al's (1975) combination of practical and written examinations, or Holmes et al's (1978) use of videotape, MCQ's and simulated patients, or Harden and Gleeson's (1979) objective structured clinical examination). Most importantly, however, such a criterion was considered inappropriate.

With regard to predictive validity of the questionnaires, an indirect indication of this may be inferred by the simple inspection of raw scores of students and registrars. For each questionnaire, the ranges of these scores do not overlap (see section 6.2.2). It may therefore be assumed that each questionnaire has a degree of predictive validity sufficient to assign group membership of a respondent. The analysis of results attempts to define this function more closely by discriminant analysis. Initially, however, calculation of the intraclass correlation co-efficient will indicate the extent to which each questionnaire discriminates between the classes of subjects in question. Appendix 6 gives a detailed discussion of the nature, purpose and formula of the intraclass

correlation co-efficient (R). Appendix 7 shows the derivation of the present value.

Calculation of the intraclass correlation co-efficient (R) indicates that both questionnaires are highly successful in discriminating between the two groups of subjects ($p < .01$ in each case). Tables 6.3 and 6.4 give summaries of the findings.

<u>Table 6.3 Intraclass Correlation Co-efficient (R) Summary of Results for Endocrinology Questionnaire</u>				
<u>Source of variation</u>	<u>SS</u>	<u>df.</u>	<u>MS</u>	<u>R</u>
Total	71621.3	69		
Between	27909.2	1	37909.2	
Within	33712.1	68	495.8	<u>0.68</u>
				($p < .01$)

<u>Table 6.4 Intraclass Correlation Co-efficient (R) Summary of Results for Neurology Questionnaire</u>				
<u>Source of variation</u>	<u>SS</u>	<u>df.</u>	<u>MS</u>	<u>R</u>
Total	50985.8	69		
Between	23442.4	1	23442.4	
Within	27543.4	68	405	<u>0.62</u>
				($p < .01$)

We may conclude, then, that both questionnaires are powerful in discriminating between the groups in question.

6.1.2 Content Validity

Although our purpose was not either to sample representatively or cover entirely the subject areas in question, content validity of the questionnaires is still a matter for consideration, since it concerns not only the adequacy of the sample but also

the appearance of the test, in other words, the test should have face validity and be, as Guilford (1954) says, "palatable to the examinee". In terms of having defined four aspects of the clinician's skill for study, and having allocated each a section of the questionnaire, adequate sampling of the skills to be studied became more amenable to control.

Initial selection of content areas was defined on the rather subjective level of the subject experts' evaluations of disorders which were either common in practice or in teaching, or important for the student and clinician to know, or both. Their evaluations, of necessity, were based on their experience of teaching at undergraduate and postgraduate level and on clinical practice in their respective fields. In addition, the content of standard textbooks was consulted. Other subject experts may possibly disagree with the selection of content made, but this would not jeopardise the usefulness of the questionnaires, since our purpose was not either to sample representatively or cover entirely the subject areas in question, but was to determine the relationship between factual knowledge, interpretation of symptoms and signs, selecting and testing diagnostic possibilities, and diagnostic acumen in those disorders selected from the population of disorders within each subject area. Selection of content, therefore, was not a crucial issue, given that it was appropriate to the respondents (students and registrars). This has been demonstrated by the process of test analysis by the intraclass correlation co-efficient and will be further demonstrated by item analysis.

Face validity of selection of MCQ options for each item, and of the questionnaire as a whole, was determined by reference to subject specialists of at least consultant status for each questionnaire separately. In total, six consultant neurologists and nine consultant physicians, specialising in endocrinology, were asked to complete the relevant questionnaire in the way indicated by placing a tick or cross as necessary after each option and to give the most likely diagnosis for each case history. Each consultant was given a clear descrip-

tion of the structure, sections, intended use and target respondents for the questionnaire.

Any options on which the subject specialists disagreed were eliminated and replaced and judged again. By this process, each questionnaire came to represent the agreed opinion of the subject specialists consulted. Face validity is thus, established.

6.1.3 Construct Validity

Since it is not our intention to interpret results on these questionnaires in terms of any psychological trait or quality, a measure of construct validity is not necessary.

6.1.4 Validity by Assumption

Guilford (1954) mentions validity by assumption, whereby it is assumed that scores measure what they are intended to measure. He gives the example of achievement test scores. In order to justify attribution of validity by assumption, it is necessary to have a clear definition of the content and of the kinds of items necessary to indicate mastery of that content. On these grounds, the content or face validity of the questionnaires, in combination with the multiple choice question and case history format may appear to justify assumptions of validity.

6.2 Validation by Test and Item Analysis

Test and item analysis procedures can be seen as a further, indirect check on the validity of a scale or questionnaire by demonstrating the goodness of its individual items and overall characteristics. Such analyses have been described as purifying the item pool (Oppeheim, 1966). In effect, such procedures occupy a grey area between validation and estimation of reliability, but in the present case we shall take them as a species of content validation.

The selection of an appropriate method of test and item analysis for the multiple choice questions presented some special problems, since the questionnaires were designed to discriminate between groups, not between individuals. As Lewy (1973)

points out, some common indices of item analysis are thus rendered inappropriate, since items which differentiate within classes of respondent will not necessarily differentiate between them. This is to be expected, since the basic units of observation (individual scores as opposed to class averages) are so different.

Parameters commonly used for selecting items for a multiple choice questionnaire to discriminate between individuals include the following:

- (a) Correlation of the item with the total test score.
- (b) Item difficulty level, customarily obtained by dividing the number of correct responses to the item by the total number of responses. A basic principle of test writing, as Wesman (1971) sees it is to "adapt the level of difficulty of the item to the group and purpose for which it is intended". In general, however, items with either a very high or very low difficulty level are discarded since items that are very easy or very difficult add little to the effectiveness of a test. An item with a difficulty level of about .5 will yield as many discriminations as possible, and therefore is most desirable. However, as Wesman (1971) continues, if the respondents are very heterogeneous, a wider spread of difficulty may be more appropriate.

It is clear that, having calculated either of the indices cited, the result could only have meaning for the particular reference group sample. As Lewy (1973) points out, two items of difficulty level at about 50 per cent and having equal correlation co-efficients with the total test score may be entirely different in the efficiency with which they discriminate among classes. The intended comparison of groups rather than individuals, rendered the three indices, as described, inappropriate. Our final method of analysis, therefore, was by test difficulty per group, and calculation of the phi co-efficient for each test item. We shall consider these in turn.

6.2.1 Test Difficulty per Group

Wood (1977) recommends calculating an item difficulty index per group to give indication of the fairness of comparing the groups. However, in a case such as ours fairness is only an appropriate concept as a matter of degree. Given that we are expecting quite large differences between our groups and that our primary interest is in the relationship between scores on the four sections of the questionnaire for students and registrars separately rather than in comparison of the absolute scores of each group, fairness in this case can only mean that the test as a whole is neither so easy that all registrars can respond correctly to all items, nor so difficult that no student can answer correctly any item. For a brief description of the calculation of test difficulty, see Appendix 8.

The results of calculating the index of test difficulty for each group separately are as follows:

Endocrinology Questionnaire

Students	- 0.18
Registrars	- 0.49

Neurology Questionnaire

Students	- 0.30
Registrars	- 0.62

The above indices may appear to attain only rather low values which would indicate that the questionnaires were difficult for the respondents, particularly for the two groups of students. However, it must be emphasised that these values are extremely conservative estimates and that the low values may, to some extent, be an artefact of the method employed. The test mean was calculated taking only those items for which the respondent had scored +5. Thus, in order for his score to be counted, the respondent must have answered correctly all five options of the item. This procedure will artificially lower the test

mean. Had the index been calculated using scores on each option of each item separately, the resultant value of the index would have been considerably higher.

However, the index as calculated is quite satisfactory for the purpose, that of showing very clearly that the difficulty levels for each questionnaire separately were different for the two groups of respondents in the expected manner, the values being relatively higher for registrars than students in each case, indicating greater test difficulty for students. In addition, the indices show that the neurology questionnaire is slightly easier than the questionnaire in endocrinology for both groups, but that the difference in difficulty levels between the two groups is similar for the two questionnaires (0.31 for the endocrinology questionnaire; 0.32 for the neurology questionnaire). The fairly wide spread of difficulty evidenced is quite appropriate in this case of very heterogeneous subjects (Wesman, 1971) and the pattern of relative difficulty is also as expected.

6.2.2. The Phi Co-efficient (ϕ)

This index is obtained by comparing the performances of the upper and lower scoring halves of a group of respondents, as described by Hubbard and Clemans (1961). Although this index is not strictly appropriate to our present case, since it is designed to identify items which successfully discriminate between individuals rather than groups of respondents, it was decided to apply it for a number of reasons. Firstly, we have no other quantitative index applied to each item separately and the phi co-efficient can provide us with such an indication of the behaviour of each item. Secondly, and conveniently, when taking the upper 25 scoring registrars, and the lower 25 scoring students for each questionnaire, their total scores do not overlap. The ranges of total scores were as follows:

Endocrinology Questionnaire:

Students	51 to 129
Registrars	141 to 178

Neurology Questionnaire:

Students 74 to 130

Registrars 143 to 175

We may therefore consider that comparing these two sets of individuals is a form of group comparison.

Calculation of phi co-efficient values was performed post hoc, for purposes of description rather than prescription; and, since the index is not thoroughly appropriate to our purpose, it was not intended that any item be excluded from the questionnaires on the basis of its associated phi co-efficient value. A value of ϕ not reaching a statistically significant level does not indicate that the item is of positive detriment to the test, but only that it does not add to its efficiency. The phi co-efficient was nonetheless considered useful as an interpretative, descriptive statistic and so was performed. Appendix 9 gives details of calculation formulae and Appendix 10 shows results for each item of the two questionnaires.

It is shown that six items from the endocrinology questionnaire, and two from the neurology questionnaire fail to reach a statistically significant value of ϕ . It must be re-emphasised that these items are not of detriment to the test, but merely fail to add to its efficiency in discriminating between the two groups. That being the case, and since values of the phi co-efficient were determined post hoc, these items are retained. The items failing to reach a statistically significant value of ϕ are not concentrated in any one section of either questionnaire.

No other indices of test validity were calculated. It is considered that satisfactory validity is demonstrated by the indices discussed. We may therefore proceed to a consideration of test reliability.

6.3 Reliability and Homogeneity of the Questionnaires

Test reliability indicates to what extent scores would be consistent across different administrations of the test to the

same sample of respondents. Where this procedure is impractical, there are alternative approaches to estimating test reliability which require only one administration of the test. The method selected here is the split-half technique, whereby a correlation co-efficient is calculated between two halves of the test. Appendix 11 gives details of the method and correction factor applied in this instance.

Results show that the values of r for each group for each questionnaire separately are statistically significant at the one per cent level, thus:

Endocrinology students	$r = 0.5597$
Endocrinology registrars	$r = 0.5004$
Neurology students	$r = 0.7179$
Neurology registrars	$r = 0.4672$

Each value of r is corrected for the length of the test, using the Spearman-Brown formula. Although these values of r are statistically significant, it must be said that they are still only moderately high and although giving support to the contention of test homogeneity, that support appears not to be very strong. However, there may be good statistical reasons for this. McNemar (1962) points out that the magnitude of the reliability co-efficient is influenced by the trait homogeneity of the sample on which it is based. His argument is worth quoting at length, since establishment of homogeneity of the questionnaires is important for the other tests and indices of validity:

"Let s_d represent the standard deviation for the restricted range, SD the standard deviation for the unrestricted range, r_{xx} the reliability for the restricted and R_{xx} the reliability for the unrestricted. If we may assume that S_e for the smaller range equals S_e for the larger range, we may write:

$$(s_d)^2 (1 - r_{xx}) = (SD)^2 (1 - R_{xx})$$

as a formula from which we can infer r_{xx} from R_{xx} and vice versa. The more homogeneous the group, xx the lower the reliability co-efficient." (p.152)

S_e here refers to the standard error of measurement. Accepting McNemar's argument, and noting the deliberate homogeneity of

groups, the otherwise only moderately high values of r become acceptable indices of test reliability and homogeneity.

6.4 Conclusions

The process of questionnaire design, construction and analysis shows instruments which display properties of both validity and reliability. Were the questionnaires to have been used for assessment or examination purposes, a different approach towards development and analysis of the instruments would have been necessary. As it is, the procedures and indices selected have shown the appropriateness of the questionnaires to the research task for which they were designed. We may now proceed to describe the statistical analyses applied to the questionnaire data.

6.5 Statistical Analyses of Data

The questionnaires in endocrinology and neurology are subjected to separate but identical statistical analyses, and patterns of results compared. When alternative statistical tests are available for a given research design, it is necessary to employ some rationale for choosing among them.

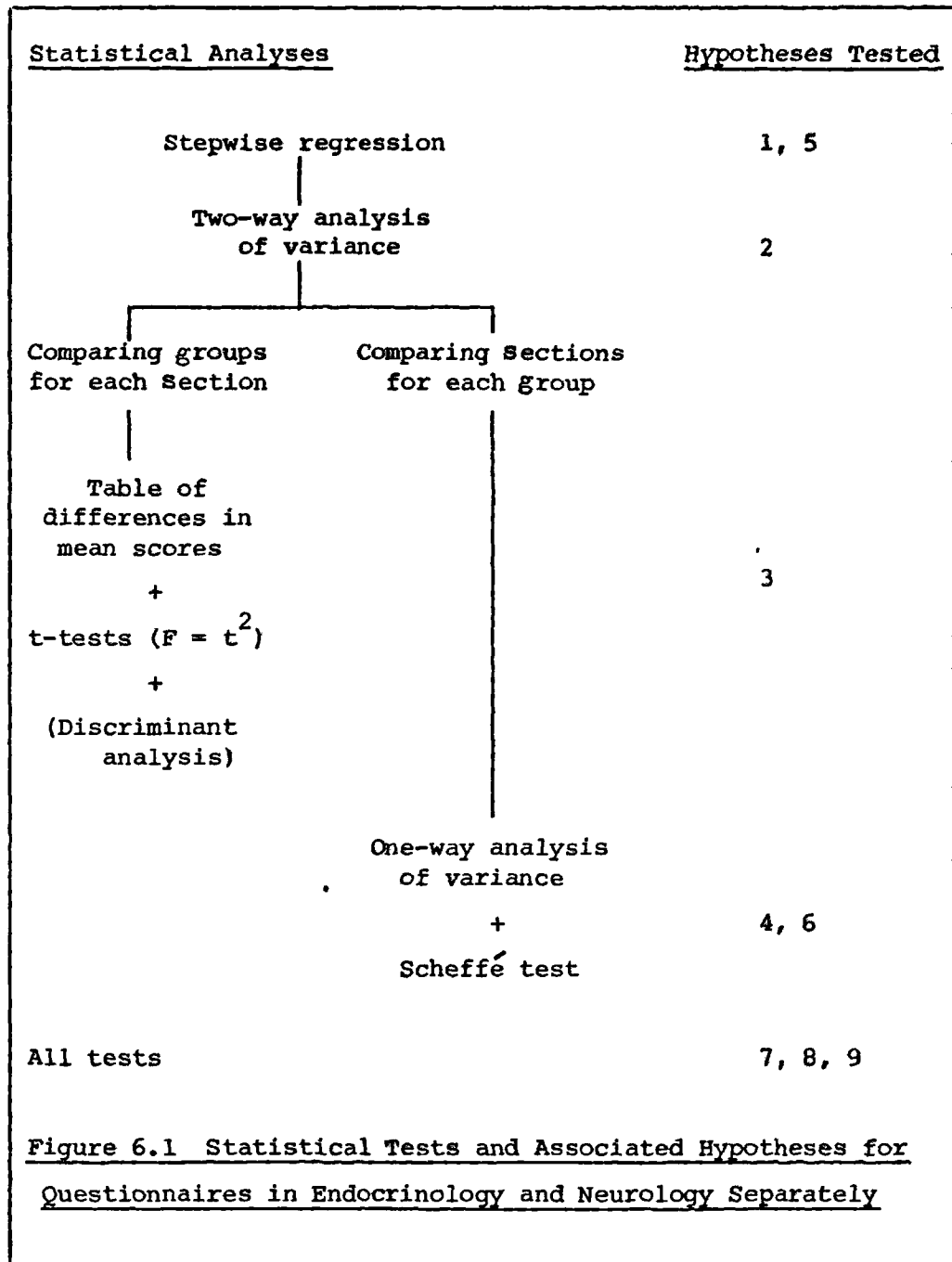
A major decision to be taken is whether to use parametric or non-parametric tests. The latter are distribution-free, the assumptions about the parent distribution being fewer in number, weaker and easier to satisfy than the assumptions underlying parametric tests. Although this is an obvious advantage, non-parametric methods are more appropriate for nominal and ordinal data, whereas parametric methods are better for interval and ratio data. In practice, non-parametric methods are often applied to data of the latter type but only use part of the information available, the data being reduced to a form such that a nominal or ordinal statistical procedure may be applied to them. Measurements are often reduced to signs or ranks, thus in data where parametric and non-parametric tests have less power, the parametric tests use more of the available information. Parametric tests were thus decided upon for this study.

However, every statistical test is only valid under certain conditions; sometimes it is possible to test whether the conditions of a particular statistical model are met, but often it is necessary to assume that they are met. The fewer these assumptions, the more generally applicable are the conclusions. It happens to be the case, however, that the most powerful tests are those which have the most extensive assumptions. The power of a test is defined as the probability of rejecting H_0 (the null hypothesis) when it is in fact false. Also to be taken into consideration are the manner in which the sample of scores was drawn and the kind of measurement or scaling employed in the operational definitions of the variables involved, that is, in the scores. Parametric tests, such as those used in this study, have a number of underlying assumptions. When these assumptions are valid these tests are the most likely of all to reject H_0 when H_0 is false.

The conditions which must be satisfied to make the parametric tests selected the most powerful are as follows (Siegel, 1956):

- a) The observations must be independent. This condition is satisfied in that the selection and score of any one subject does not bias the selection and score of any other.
- b) The observations must be drawn from normally distributed populations. Each sample is taken from the general populations of final year clinical medical students and medical registrars. There is no reason to assume that these populations are other than normally distributed for the variables tested.
- c) These populations must have the same variances. Although this is not tested, we have no reason to assume that this is other than the case.
- d) The variables involved must have been measured in at least an interval scale. This condition is satisfied.

We may now consider the statistical analyses applied to test the associated Hypotheses 1 to 9 (see section 5.1). The tests, their inter-relationship and associated hypotheses are summar-



ried in Figure 6.1. The order of tests may be seen as representing a progressive specificity of analysis. We may now consider each test in turn.

6.5.1 Stepwise Linear Regression

Hypotheses 1 and 5 concern the relationship between scores on section D and scores on the other three sections. In other words, the relationship between the skills tested in sections

A, B and C and skill in actually making a diagnosis is analysed. Free stepwise linear regression analysis was used to determine the relative importance of each predictor variable (sections A, B and C) in explaining the variance associated with the criterion variable (section D). This analysis was performed using the SPSS sub-programme 'Regression' on the ULCC CDC6600 computer. The programme selects variables in the order in which they best account for the variation of the criterion variable, based on the reduction of the error sum of squares. The variable which makes the greatest reduction in the error sum of squares is entered next into the regression equation at each step.

Regression analysis, then, is a general statistical technique whereby one can analyse the relationship between a dependent or criterion variable and a set of independent or predictor variables. The technique may be used either as a descriptive tool by which the linear dependence of one variable on others is summarised and analysed, or as an inferential tool by which the relationships existing in the population are evaluated from the examination of sample data. The sample size of the present study is too small to admit of strong inference, and so this analysis is primarily descriptive. We may, therefore, find the best linear prediction equation and evaluate its predictive accuracy. The proper interpretation of regression summary tables requires recourse to the associate correlation matrix. For short notes on this and other points in the interpretation of regression summary tables, see Appendix 12.

6.5.2 Two-Way Analysis of Variance

Hypothesis 2 concerns the reciprocal effects of groups and sections. Two-way analysis of variance is appropriate when classification of subjects or scores into groups is made on the basis of two or more variables - in this instance, student or registrar and sections A, B, C and D. The resultant analysis of variance is a method for dividing the variation observed in the data into different parts, each part assignable to a known source or factor. The relative magnitude of

variation resulting from different sources may be assessed to ascertain whether a particular part of the variation is greater than expectation under the null hypothesis. The analysis of variance is thus inextricably associated with the design of research. The analysis was performed by hand, using a Casio fx-39 Scientific Calculator.

Two-way analysis of variance yields three values of F (the ratio of two sample variances or variance estimates) testing the significance of row (F_r), column (F_c) and interaction (F_i) effects. In this case, rows are associated with groups of subjects (students and registrars) while columns are associated with sections on the questionnaire. The design, therefore, is of two rows by four columns with 35 observations in each cell. The third value of F is the interaction term (F_i). Where the value of F_i is statistically significant, the interpretation of the main effects observed must be with some qualification (McNemar, 1962), the interaction term being a measure of the extent to which the criterion mean for the combination ab_{ij} cannot be predicted from the sum of the corresponding main effects (Winer, 1962). Our main focus of attention, therefore, is on the interaction term.

6.5.3 Table of Differences between Groups in Mean Scores per Section with t-tests. Supported by Discriminant Analysis

This analysis provides a deeper understanding of the results of the two-way analysis of variance value of F_r , when applied for each section of each questionnaire separately. In this instance the value of t is calculated from a one-way analysis of variance value of F calculated for the two groups ($F = t^2$). The statistic does not require the initial stages of calculation, since values derived in the calculation of the two-way analysis of variance may be borrowed.

Findings from these tests must be interpreted in the light of the different difficulty levels of the test for each group. The difficulty levels as derived (section 6.2.1) would imply statistically significant differences between students and

registrars on all sections of each questionnaire. However, the magnitude of those differences, as shown in a table comparing observed differences between students and registrars in mean scores on each section, may be of greater interpretative value. This analysis yields results relevant to Hypothesis 3.

Supporting this analysis is a discriminant analysis which identifies the extent and nature of the power of the sections of each questionnaire to differentiate or discriminate between the two groups of subjects. Appendix 13 discusses this test in detail.

6.5.4 One-Way Analysis of Variance with Scheffé Test, Comparing Scores on Sections for Students and Registrars Separately

This analysis provides a deeper understanding of the results of the two-way analysis of variance value of F_c , when applied for students and registrars separately. The statistic does not require the initial stages of calculation, since values derived in the calculation of the two-way analysis of variance may be borrowed. Application of the Scheffé method following the F test allows a complete comparison of pairs of means; that is, comparing each section with every other section one at a time. The differences between some pairs of means may be statistically significant, while other differences may not be so. See Appendix 14 for a description of the Scheffé method as applied here. This analysis yields results relevant to Hypotheses 4 and 6.

Note: For all analyses of variance data were normalised, since the differing ranges of the theoretical maximum and minimum raw scores preclude the necessary homogeneity of within group variance. See Appendix 15 for a description of normalisation procedures.

6.6 Summary

A description of the structure and content of the questionnaires in endocrinology and neurology is given with the number of questions and ranges of raw scores possible. The validity of the

questionnaires is discussed in terms of criterion oriented validity, content validity, construct validity, validity by assumption and by test and item analysis. Statistical procedures are reported. The reliability and homogeneity of the questionnaires are established. All statistical analyses of questionnaire data are described and discussed and associated research hypotheses identified.

CHAPTER SEVEN

The Questionnaire Study 2. Results

This chapter presents only results per se of the statistical analyses performed. These will be presented in relation to the separate analyses applied as described in section 6.5 above. Full interpretation and discussion of the results in relation to the research hypotheses stated in section 5.1 above will follow in Chapter Eight.

7.1 Stepwise Linear Regression

Four analyses by stepwise regression were performed, for students and registrars separately, for endocrinology and neurology, also separately. Before reporting the results of each of these, an initial perusal and discussion of the associated correlation matrices is advisable, in order to assist accurate interpretation of the stepwise regression values of R^2 .

- Tables 7.1, 7.2, 7.3 and 7.4 show the correlation matrices between all variables for the four stepwise linear regression analyses. It will be noted that registrars demonstrate fewer statistically significant correlations than do students. Table 7.5 summarises the relative incidence of statistically significant correlation co-efficients across groups of subjects.

Table 7.1 Correlation Matrix Between all
Variables - Endocrinology Students

<u>Variable</u>	<u>Section A</u>	<u>Section B</u>	<u>Section C</u>	<u>Section D</u>
Section A	1.0000	<u>0.7292</u>	0.2641	<u>0.4558</u>
Section B		1.0000	<u>0.4377</u>	<u>0.6162</u>
Section C			1.0000	<u>0.4269</u>
Section D				1.0000

Underlined values are statistically significant ($p < .01$ df.33)

Table 7.2 Correlation Matrix Between all
Variables - Endocrinology Registrars

<u>Variable</u>	<u>Section A</u>	<u>Section B</u>	<u>Section C</u>	<u>Section D</u>
Section A	1.0000	0.4121	-0.1328	-0.0830
Section B		1.0000	0.0504	0.1045
Section C			1.0000	-0.0671
Section D				1.0000

Underlined values are statistically significant ($p < .01$ df.33)

Table 7.3 Correlation Matrix Between all
Variables - Neurology Students

<u>Variable</u>	<u>Section A</u>	<u>Section B</u>	<u>Section C</u>	<u>Section D</u>
Section A	1.0000	<u>0.7457</u>	0.3751	0.4056
Section B		1.0000	<u>0.6438</u>	<u>0.5125</u>
Section C			1.0000	0.3703
Section D				1.0000

Underlined values are statistically significant ($p < .01$ df.33)

Table 7.4 Correlation Matrix Between all
Variables - Neurology Registrars

<u>Variable</u>	<u>Section A</u>	<u>Section B</u>	<u>Section C</u>	<u>Section D</u>
Section A	1.0000	<u>0.6626</u>	0.2568	0.1812
Section B		1.0000	0.3125	<u>0.4839</u>
Section C			1.0000	0.1557
Section D				1.0000

Underlined values are statistically significant ($p < .01$ df.33)

Table 7.5 Relative Incidence of Statistically
Significant Correlation Co-efficients Across
Groups of Subjects

<u>Sections (Variables)</u>	<u>Endocrinology</u>		<u>Neurology</u>	
	<u>Students</u>	<u>Registrars</u>	<u>Students</u>	<u>Registrars</u>
A and B	*		*	*
A and C				
A and D	*			
B and C	*		*	
B and D	*		*	*
C and D	*			

Registrars show no statistically significant correlations between sections of the endocrinology questionnaire. The patterns of statistically significant correlations are more alike for students and registrars in neurology than in endocrinology. In particular, it is interesting that for both students and registrars section B correlates statistically significantly with section D for neurology, whereas, for endocrinology, all sections correlate statistically significantly with section D for students and no section does for registrars.

Having access to the relevant correlation matrices, we may now proceed to the results of the four stepwise linear regression analyses, with section D as the dependent (criterion) variable.

7.1.1 Endocrinology Students

Table 7.6 presents the stepwise regression analysis summary table. The predictor variables together account for 41 per cent of the variance on the criterion. We see that the first variable entered into the regression equation is section B (interpretation of symptoms and signs). This variable explains almost 38 per cent of the variance on the criterion variable ($R^2 = 0.3797$). Table 7.1 shows, as expected, that the correlation between sections B and D is statistically significant

Table 7.6 Summary of Free Stepwise Regression
Analysis - Endocrinology Students

<u>Variable Entered</u>	<u>Multiple R</u>	<u>Multiple R²</u>	<u>Increase in R²</u>
Section B	0.6162	0.3797	0.3797
Section C	0.6405	0.4102	0.0305
Section A	0.6410	0.4109	0.0007

($r = 0.4558$). The next variable to be entered into the equation is section C, but this accounts for only about 3 per cent more of the variance on section D (increase in $R^2 = 0.0305$); section A is entered next and repeats this pattern (increase in $R^2 = 0.0007$). Since both sections C and A are statistically significantly correlated with section B ($r = 0.4377$ and $r = 0.7292$, respectively), they would be expected to add little predictive power to the regression equation. We may say, then, that section B is the best predictor of scores on section D for students in endocrinology, but unless we consider the predictive power of each variable separately (that is, as if each were entered first into the regression equation), we cannot determine how much better is this variable than the others. Table 7.7 shows the values of multiple R and multiple R^2 for each of the predictor variables when entered first into the regression equation, thereby exerting its entire predictive power.

Table 7.7 Values of R and R² for Each Predictor
Variable Independently of all Others -
Endocrinology Students

<u>Variable</u>	<u>Multiple R</u>	<u>Multiple R²</u>
Section A	0.4680	0.2190
Section B	0.6162	0.3797
Section C	0.3679	0.1353

Section A, independently of section B, would account for 22 per cent of the variance on section D ($R^2 = 0.2190$), whereas section C would account for only 13 per cent ($R^2 = 0.1353$) and so has least predictive value. Since section A is correlated more highly with section D than is section C, these are the expected relative predictive powers.

7.1.2 Endocrinology Registrars

Table 7.8 presents the stepwise regression analysis summary table. Again, the first variable entered into the regression equation is section B. However, this explains only about one per cent of the variance on the criterion variable ($R^2 = 0.0109$). Section A is entered next, accounting for a further two per cent of the variance (increase in $R^2 = 0.0192$), followed by section C which adds only a negligible amount of predictive power (increase in $R^2 = 0.0095$). Table 7.2 shows no statistically significant correlations between any two variables (criterion or predictors), the lack of predictive power here demonstrated is therefore to be expected. The magnitude of that lack, however, may seem surprising. The predictor variables together account for only four per cent of the variance on the criterion. Lack of any statistically significant correlation coefficients obviates any possible need to refer to the predictive power of each predictor independent of the others.

<u>Table 7.8 Summary of Free Stepwise</u> <u>Regression Analysis - Endocrinology Registrars</u>			
<u>Variable entered</u>	<u>Multiple R</u>	<u>Multiple R^2</u>	<u>Increase in R^2</u>
Section B	0.1045	0.0109	0.0191
Section A	0.1734	0.0301	0.0192
Section C	0.1988	0.0395	0.0095

7.1.3 Neurology Students

Table 7.9 shows that the first variable entered into the regression equation is, as for students and registrars in endocrinology, section B. This variable explains 26 per cent of the variance on section D ($R^2 = 0.2627$). As expected, Table 7.3 shows a statistically significant correlation between sections B and D ($r=0.5125$). The next two variables entered account for only a further 0.5 per cent of the variance on the criterion (increase in $R^2 = 0.0050$), accounting for a total of almost 27 per cent ($R^2 = 0.2677$). Since both sections C and A correlate statistically significantly with section B ($r = 0.6438$ and 0.7457 , respectively), we would not expect them to add greatly to the value of R^2 established by section B. In this case, it is worth considering the predictive power of these variables independently of section B.

Table 7.9 Summary of Free Stepwise Regression
Analysis - Neurology Students

<u>Variable entered</u>	<u>Multiple R</u>	<u>Multiple R^2</u>	<u>Increase in R^2</u>
Section B	0.5125	0.2627	0.2627
Section C	0.5152	0.2654	0.0027
Section A	0.5174	0.2677	0.0023

Section C, independently of section B, would account for 16 per cent of the variance on section D ($R^2 = 0.1618$), whereas section A would account for only 13 per cent ($R^2 = 0.1362$) and so has least predictive value. However, these are relatively small values, and the entire set of variables accounts for only slightly more than a quarter of the variance on the criterion. In endocrinology, these predictors account for rather more than one third of the variance on the criterion.

Table 7.10 Values of R and R^2 for Each Predictor
Variable Independently of all Others -
Neurology Students

<u>Variable</u>	<u>Multiple R</u>	<u>Multiple R^2</u>
Section A	0.3691	0.1362
Section B	0.5125	0.2627
Section C	0.4023	0.1618

7.1.4 Neurology Registrars

Table 7.11 shows that the first variable entered into the regression equation is, as for all previous analyses, section B which explains 23 per cent of the variance on the criterion ($R^2 = 0.2341$). The next variable entered, section A, accounts for only a further three per cent (increase in $R^2 = 0.0347$), while the additional predictive value of section C is negligible (increase in $R^2 = 0.0003$), making a total of 27 per cent ($R^2 = 0.2691$).

Table 7.4 shows that sections A and B correlate statistically significantly, whereas section C does not correlate statistically significantly with any other variable. Nonetheless, section A is entered second into the regression equation. It may, therefore, be useful to consider the predictive value of each predictor variable independent of all others.

Table 7.11 Summary of Free Stepwise Regression
Analysis - Neurology Registrars

<u>Variable entered</u>	<u>Multiple R</u>	<u>Multiple R^2</u>	<u>Increase in R^2</u>
Section B	0.4839	0.2341	0.2341
Section A	0.5185	0.2688	0.0347
Section C	0.5188	0.2691	0.0003

Table 7.12 shows that section C has very little predictive power ($R^2 = 0.0222$), whereas section A accounts for 17 per cent of the variance on the criterion ($R^2 = 0.1740$). However, these, again, are relatively small values.

<u>Table 7.12 Values of R and R^2 for Each Predictor</u> <u>Variable Independently of all Others -</u> <u>Neurology Registrars</u>		
<u>Variable</u>	<u>Multiple R</u>	<u>Multiple R^2</u>
Section A	0.4171	0.1740
Section B	0.4839	0.2341
Section C	0.1489	0.0222

7.1.5 Comparison of Groups

Results of stepwise linear regression analyses are similar across all groups and particularly across registrars and students separately in terms of the relative contributions of the predictor variables. Figure 7.1 illustrates this point. For all groups of subjects section B has most predictive value for scores on section D, but differs in the extent of this predictive power. For students and registrars in neurology, the values are similar, 26 per cent and 23 per cent respectively. For students and registrars in endocrinology, the values are very different, being higher for students and much lower for registrars (38 per cent and one per cent, respectively).

For all groups, sections A and C add almost no predictive power to that attained by section B. The additional predictive value varies between 3.5 per cent (neurology registrars) and 0.5 per cent (neurology students). Independently of section B, section A has some predictive value for endocrinology students (22 per cent), neurology students (13 per cent) and neurology registrars (17 per cent). Section C has some predictive value for endocrinology students (13 per cent) and neurology students (16 per cent) and a negligible amount for neurology registrars

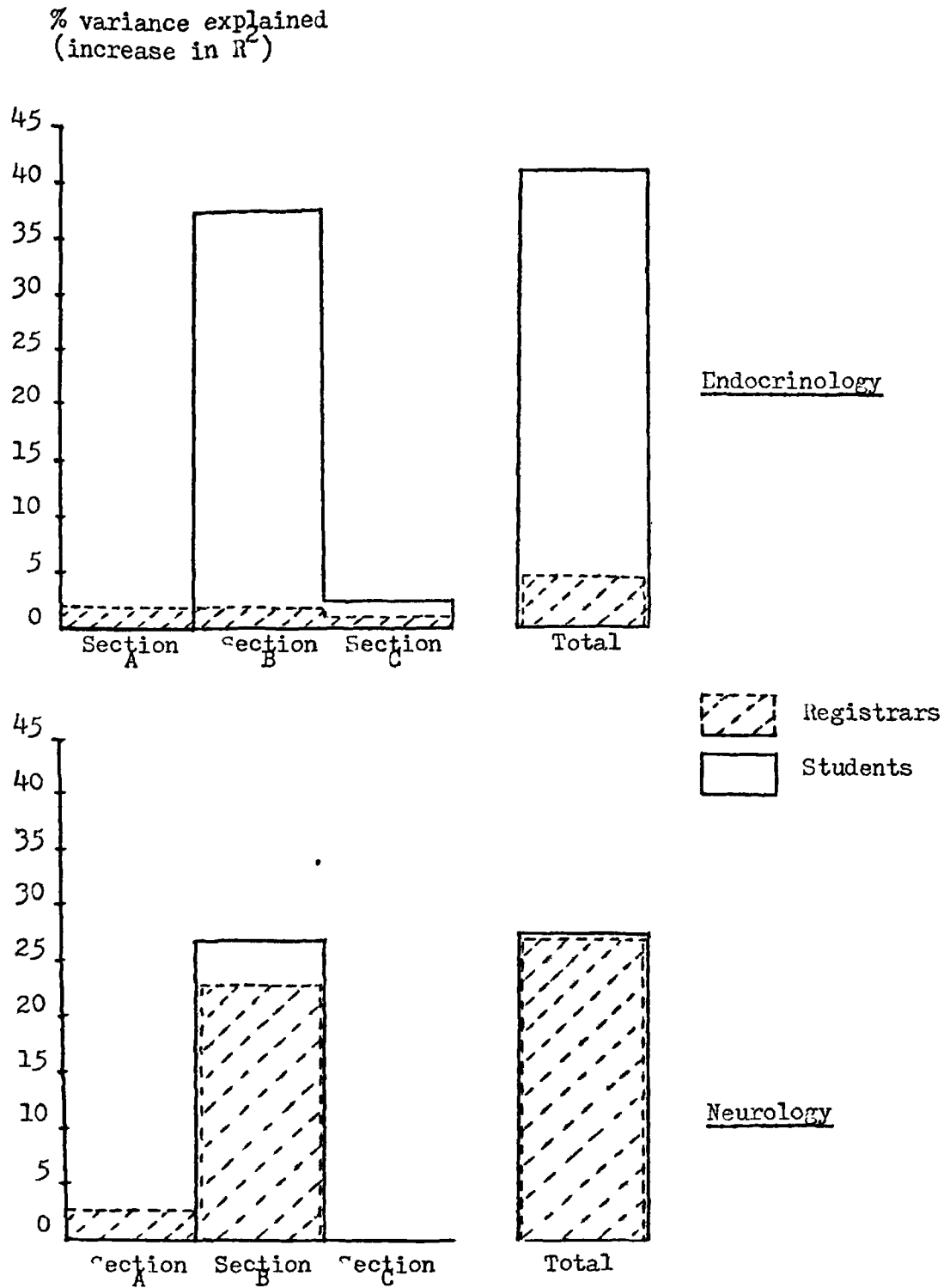


Figure 7.1 Diagrammatic Representation of Relative Contributions of Section A, B and C in Predicting Scores on Section D (Free Stepwise Regression)

(two per cent).

In terms of the total predictive value of the set of predictor variables, the results for endocrinology and neurology are dissimilar. In turn, the relative results for students and registrars differ in endocrinology and neurology, being almost identical for groups in neurology (27 per cent; $R^2 = 0.2677$ for students and $R^2 = 0.2691$ for registrars), and dissimilar for groups in endocrinology, students' scores having greater predictive power (41 per cent, $R^2 = 0.4109$) and registrars' having less (four per cent, $R^2 = 0.0395$).

The only remaining point of comparison is seen in the correlation matrices, which reflect and give greater explanatory power to the results of the regression analyses. In endocrinology, the incidence of statistically significant correlation coefficients for students and registrars is very different. For students, all variables correlate statistically significantly with all others, with the exception of sections A and C. For registrars, the picture is entirely different, showing no statistically significant correlations between any variables.

In neurology, the patterns of statistically significant correlation coefficients are similar for students and registrars (section A with section B, and section B with section D), while students demonstrate an additional statistically significant correlation between sections B and C.

7.2 Two-way Analysis of Variance

Two-way analysis of variance was performed separately for endocrinology and neurology in order to determine, in particular, the interaction effects between groups and sections. In addition, values of F for rows and columns (groups and sections, respectively) are obtained.

7.2.1 Endocrinology

Table 7.13 shows the two-way analysis of variance summary table. We may observe the interaction term is not statistically significant, while the main effects of row and column both

reach statistically significant values. The values of F_r and F_c will be afforded further meaning by the analyses described in sections 7.3 and 7.4. The statistically non-significant interaction term ($F_i = 3.0920$) suggests that, in variance terms, the difference between groups is constant across all sections. The values of F_r and F_c indicate statistically significant differences between groups and sections, respectively.

<u>Table 7.13 Two-Way Analysis of Variance Summary</u>			
<u>Table - Endocrinology</u>			
<u>Source of variation</u>	<u>Sums of squares</u>	<u>Df</u>	<u>Variance estimate</u>
Rows (groups)	11,689	1	11,689
Columns (sections)	1,846	3	615
Interaction	1,230	3	410
Within	36,069	272	133
Total	50,834	279	
$F_i = 3.0920$	$F_r = 88.1540$	$F_c = 4.641$	
(df 3,272)	(df 1,272)	(df 3,272)	
NS	$p < .01$	$p < .01$	

7.2.2 Neurology

Table 7.14 indicates a statistically significant interaction term ($F_i = 7.7080$), as well as statistically significant values of F_r ($F_r = 97.6302$) and F_c ($F_c = 18.5424$). These latter two values are further discussed in sections 7.3 and 7.4. The statistically significant interaction term makes precarious the interpretation of the nature of the row and column differences evidenced. Recourse to further analyses, therefore, is helpful. The present analysis, however, shows that there is an interaction between subjects and skills. In other words, we may say that the magnitude of difference, in variance terms, between

Table 7.14 Two-Way Analysis of Variance Summary

Table - Neurology

<u>Source of variation</u>	<u>Sums of squares</u>	<u>Df</u>	<u>Variance estimate</u>
Rows (groups)	12,236	1	12,236
Columns (sections)	6,972	3	2,324
Interaction	2,898	3	966
Within	34,089	272	125
Total	56,195	279	
$F_i = 7.7080$ (df 3,272) $p < .01$	$F_r = 97.6302$ (df 1,272) $p < .01$	$F_c = 18.5424$ (df 3,272) $p < .01$	

students and registrars is dependent upon the section (or skill) being tested, and is not uniformly different as is the case with endocrinology. The statistically significant main effects, however, do indicate differences in scores between the group of students and the group of registrars, as well as overall differences in the scores on sections. We may now proceed to elucidate further these differences.

7.3 One-Way Analysis of Variance, with Table of Observed Differences in Mean Scores, Comparing Students' and Registrars' Scores on Each Section Separately

This analysis provides explanatory information concerning the two-way analysis of variance value of F_r , when applied for each section of each questionnaire separately. Of course, the difficulty levels given in section 6.2.1 would cause us to expect statistically significant differences between students' and registrars' scores on all sections of both questionnaires. It is therefore almost a routine exercise to demonstrate this. However, the associated observed differences found between students and registrars on each section of each questionnaire

may prove more revealing and interesting. A statistically significant level of difference, after all, tells us nothing about the size of that difference. As Hays (1963) indicates, "the actual difference obtained is always the best estimate ... of the true difference between the population means". In other words, difference may be statistical, or may have a meaning which is significant educationally or psychologically.

7.3.1 Endocrinology

Table 7.15 presents the values of mean scores per section for students and registrars, and the associated value of $t (\sqrt{F})$. It will be noted that all differences between mean scores, as expected, are statistically significant.

<u>Table 7.15 Observed Differences in Mean Scores per Section between Students and Registrars - Endocrinology</u>				
	Section A	Section B	Section C	Section D
<u>Students</u>				
Raw mean score	37.48	32.48	30.60	4.57
Normalised mean score	81.24	76.51	69.14	76.40
<u>Registrars</u>				
Raw mean score	46.31	44.68	56.11	5.51
Normalised mean score	88.59	87.23	87.40	91.74
<u>Observed difference</u>				
Raw scores	8.83	12.20	25.51	0.94
Normalised scores	7.35	10.72	18.26	15.34
F value	28.41	81.45	57.83	12.53
t value	5.33	9.02	7.58	3.54
df (n-2)	68	68	68	68
	(p<.01)	(p<.01)	(p<.01)	(p<.01)

Registrars have higher scores than students on all sections. Comparing normalised scores, we see that the greatest difference is on section C, closely followed by section D. The least difference is on section A.

7.3.2 Neurology

Table 7.16 presents the values of mean scores per section for students and registrars, and the associated values of t and F. Again, as expected, all differences are statistically significant, registrars scoring more highly on all sections.

Comparing normalised scores, the greatest difference is on section D (24.29). This is much larger than the next largest value of 10.55 on section A. Indeed, there is very little difference (less than two per cent) between the values of the differences between means on sections A, B and C.

<u>Table 7.16 Observed Differences in Mean Scores per Section between Students and Registrars - Neurology</u>				
	Section A	Section B	Section C	Section D
<u>Students</u>				
Raw mean score	37.03	31.57	45.91	9.06
Normalised mean score	79.42	81.57	80.56	61.65
<u>Registrars</u>				
Raw mean score	50.88	40.17	60.03	12.94
Normalised mean score	89.97	90.17	90.01	85.94
<u>Observed difference</u>				
Raw scores	13.85	8.60	14.12	3.88
Normalised scores	10.55	8.60	9.45	24.29
F value	17.62	25.80	29.07	34.94
t value	4.20	5.08	5.39	5.91
df (n-2)	68	68	68	68
	(p<.01)	(p<.01)	(p<.01)	(p<.01)

7.3.3 Comparison of Endocrinology and Neurology

Firstly, as expected, for both endocrinology and neurology, differences between scores for students and registrars are statistically significant for every section, registrars' scores attaining higher values than those of students.

Secondly, we may consider the relative values of observed differ-

ences for each section. Here, we see considerable differences between endocrinology and neurology. For endocrinology, section C shows the greatest difference between mean scores, whereas for neurology, section D does. The least difference is evidenced by sections A and B for endocrinology and neurology respectively. Table 7.17 summarises these comparisons, as well as demonstrating the ratios of differences in mean scores from section to section. It is clearly shown that, for neurology, section D is pre-eminent in differentiating between scores, with sections A, B and C showing very similar degrees of difference. For endocrinology, however, both section C and section D show large differences in mean scores, although even the section C difference does not reach the magnitude of difference displayed by section D in neurology.

Table 7.17 Relative Magnitude and Ratios of Observed Differences between Students and Registrars on Mean Scores per Section for Endocrinology and Neurology

Section in order of magnitude of observed difference		Value of observed difference		Ratio of values of observed differences	
<u>Endo- crinology</u>	<u>Neurology</u>	<u>Endo- crinology</u>	<u>Neurology</u>	<u>Endo- crinology</u>	<u>Neurology</u>
C	D	18.26	24.29	2.48	2.82
D	A	15.34	10.55	2.09	1.23
B	C	10.72	9.45	1.46	1.10
A	B	7.35	8.60	1.00	1.00

It may be held that these observed differences are either unreliable or invalid because of the different difficulty levels of the tests for students and registrars, as shown in section 6.2.1 above. However, it will be noticed that the relative difficulty levels for students and registrars are comparable across endocrinology and neurology questionnaires. It is, therefore, permissible to make comparisons of the present nature, and to

maintain that the student-registrar differences measured are not constant across the two specialities.

Results of discriminant analysis prove unhelpful in further elucidating these points. The different purpose of the analysis yields slightly different results. However, since they are interesting in their own right, they are reported and discussed in Appendix 16.

7.4 One-Way Analysis of Variance with Scheffé Test, Comparing Scores on Sections for Students and Registrars Separately

This analysis, when performed separately for students and registrars, provides further information about the derivation of the statistically significant values of F_c shown by the two-way analysis of variance for both endocrinology and neurology. From each of those analyses, we may only say that, when scores for students and registrars are taken together, there are statistically significant differences between scores on the four sections of each questionnaire. However, we cannot say how those sections interrelate for each group of subjects separately, or whether or not they relate in the same way for students and registrars and for endocrinology and neurology. This analysis, then, was performed for endocrinology and neurology with students and registrars separately, in order to understand more fully the meaning of the statistically significant F_c values.

7.4.1 Endocrinology Students

Table 7.18 shows the one-way analysis of variance summary table and the resulting statistically significant value of F ($F = 4.68$, $df\ 3,136$). We may conclude that the skill (section) tested affects the score gained and proceed to analyse the manner in which scores differ across sections by applying the Scheffé test for comparison of pairs of means (see Appendix 14).

Table 7.19 shows that only one statistically significant difference emerges, this being between sections A and C. This

Table 7.18 One-Way Analysis of Variance Summary
Table for Scores on Sections - Endocrinology Students

<u>Source of Variation</u>	<u>Sums of Squares</u>	<u>Df</u>	<u>Variance estimate</u>
Between	2,617	3	872
Within	25,309	136	186
Total	27,926	139	

F = 4.688 (df 3,136)
p < .01

could have been predicted from an inspection of Table 7.1. This value of F (F = 13.7450, df 3,136) arises from comparing the section with the highest mean score, with the section with the lowest mean score. No other statistically significant differences between sections are found.

Table 7.19 Results of Scheffé Method of Complete Set of Comparisons - Endocrinology Students

<u>Sections Compared</u>	<u>Difference between Means</u>	<u>F Value</u>	<u>Level of significance</u>
A, B	4.73	2.0949	NS
A, C	12.10	13.7450	p < .10; p < .01
A, D	4.84	2.1937	NS
B, C	7.37	5.1033	NS
B, D	0.11	0.0011	NS
C, D	7.26	4.9521	NS

F' = 11.79 (p < .01); 6.39 (p < .10) df 3,136

Section means: A = 81.24; B = 76.51; C = 69.14; D = 76.40

Standard deviations: A = 6.54; B = 10.07; C = 10.40; D = 22.32

7.4.2 Endocrinology Registrars

Table 7.20 shows no statistically significant difference between scores for sections for registrars in endocrinology. This finding is also congruent with Table 7.2. Table 7.21 confirms this finding, demonstrating very small observed differences in mean scores across all pairs of sections.

Table 7.20 One-Way Analysis of Variance Summary

Table for Scores on Sections -

Endocrinology Registrars

<u>Source of Variation</u>	<u>Sums of Squares</u>	<u>Df</u>	<u>Variance Estimate</u>
Between	433	3	144
Within	10,877	136	80
Total	11,310	139	

F = 1.8000 (df 3,136)
NS

Table 7.21 Results of Scheffé Method of Complete Set of Comparisons - Endocrinology Registrars

<u>Sections compared</u>	<u>Difference between means</u>	<u>F value</u>	<u>Level of significance</u>
A, B	1.36	0.4047	NS
A, C	1.19	0.3098	NS
A, D	3.15	2.1711	NS
B, C	0.17	0.0063	NS
B, D	4.51	4.4505	NS
C, D	4.34	4.1213	NS

F' = 11.79 (p<.01); 6.39 (p<.10) df 3,136

Section means: A = 88.59; B = 87.23; C = 87.4; D = 91.74

Standard deviations: A = 5.00; B = 5.95; C = 9.88; D = 12.62

7.4.3 Neurology Students

Table 7.22 shows a statistically significant value of F ($F = 17.5485$, $df\ 3,136$), indicating that skills tested affect scores gained. Table 7.23 adds to this result by showing that there are statistically significant differences between section D and all other sections. In this case, the score on section D is statistically significantly lower than all other scores.

Table 7.22 One-Way Analysis of Variance Summary Table for
Scores on Sections - Neurology Students

<u>Source of Variation</u>	<u>Sums of Squares</u>	<u>Df</u>	<u>Variance estimate</u>
Between	9,319	3	3,106
Within	24,073	136	177
Total	33,392	139	

$F = 17.5480\ (df\ 3,136)$
 $p < .01$

Table 7.23 Results of Scheffé Method of Complete Set of
Comparisons - Neurology Students

<u>Section compared</u>	<u>Difference between means</u>	<u>F value</u>	<u>Level of significance</u>
A, B	2.15	0.4570	NS
A, C	1.14	0.1285	NS
A, D	17.77	31.2189	$p < .10$; $p < .01$
B, C	1.01	0.1108	NS
B, D	19.92	39.2303	$p < .10$; $p < .01$
C, D	18.91	35.3529	$p < .10$; $p < .01$

$F' = 11.79\ (p < .01)$; $6.39\ (p < .10)\ df\ 3,136$

Section means: A = 79.42; B = 81.57; C = 80.56; D = 61.65

Standard deviations: A = 12.55; B = 7.45; C = 6.78; D = 21.10

7.4.4 Neurology Registrars

Table 7.24 shows no statistically significant difference between scores for sections, implying that the skill tested does not affect the score gained. Table 7.25 confirms this finding, demonstrating very small observed differences in mean scores across all pairs of sections.

<u>Table 7.24 One-Way Analysis of Variance Summary Table</u> <u>for Scores on Sections - Neurology Registrars</u>			
<u>Source of Variation</u>	<u>Sums of Squares</u>	<u>Df</u>	<u>Variance estimate</u>
Between	444	3	148
Within	10,287	136	74
Total	10,730	139	
F = 2.000 (df 3,136)			
NS			

<u>Table 7.25 Results of Scheffé Method of Complete Set of</u> <u>Comparisons - Neurology Registrars</u>			
<u>Sections Compared</u>	<u>Difference between means</u>	<u>F Value</u>	<u>Level of significance</u>
A, B	0.2	0.0092	NS
A, C	0.04	0.0004	NS
A, D	4.03	3.7576	NS
B, C	0.16	0.0059	NS
B, D	4.23	4.1399	NS
C, D	4.07	3.8326	NS
F' = 11.79 (p<.01); 6.39 (p<.10) df 1,136			
Section means: A = 89.97; B = 90.17; C = 90.01; D = 85.94			
Standard deviations: A = 7.99; B = 6.71; C = 7.09; D = 12.06			

7.4.5 Comparison of Groups

Considering students and registrars, clear similarities and differences arise, in that both groups of students have associated statistically significant values of F , whereas neither group of registrars demonstrates any such statistically significant values. We may conclude that students' scores differ between sections more than do those of registrars.

Inspection of the differences between means confirms this interpretation, as does calculation of the ranges of means across sections (endocrinology students, 12.1; neurology students, 19.92; endocrinology registrars, 4.51; neurology registrars, 4.23).

Looking more deeply at the nature of the statistically significant values of F for endocrinology students ($F = 4.688$; $df\ 3,136$) and neurology students ($F = 17.5480$; $df\ 3,136$) by considering the results of comparisons of pairs of means, it is clear that the two groups are actually dissimilar. For neurology students, the value of F seems largely to be related to low scores on section D only. Inspection of the differences between other pairs of means (Table 7.23) shows very low values of 2.15, 1.14 and 1.01. For students in endocrinology, the picture is quite different. No one section scores consistently, statistically significantly differently from all other sections. Indeed, the only statistically significant difference occurs between sections at either extreme of the range of mean scores; section A scores most highly (mean = 81.24) while section C has the lowest mean score (mean = 69.14).

Considering the non-statistically significant results of the one-way analysis of variance for registrars in endocrinology and neurology, it is interesting to note that the relative values of the differences between means (Tables 7.21 and 7.25) show the same pattern in endocrinology and neurology, in that the greatest differences are associated with section D in both cases. However, this is due to the relatively high mean

score on section D for endocrinology registrars, and the relatively low mean score on section D for neurology registrars. Standard deviations on these scores are similar (12.62 for endocrinology; 12.06 for neurology). This information provides only a vague indication of a possible real difference between the groups. The lack of any statistical significance attached to these values precludes strong inference or generalisation of the observations.

Comparing results on endocrinology and neurology, we find a greater differentiation of scores between sections for students than for registrars in both cases, although the nature of those differences varies in the manner already discussed.

7.5 Summary

Chapter Seven presents results of the main statistical analyses of data from the questionnaires in endocrinology and neurology. Results of the discriminant analysis prove non-contributory and are dealt with separately in Appendix 16. Results reported here are from analysis by stepwise linear regression interpreted in the light of full correlation matrices for each group across all questionnaire sections, and independent values of R and R^2 for each predictor variable. This is followed by results from two-way analysis of variance, elucidated by means of one-way analysis of variance and tables of observed differences in mean scores per section between students and registrars. One-way analysis of variance, with Scheffé tests is also used to compare scores across sections for each group of subjects separately. All results are reported for neurology and endocrinology separately, after which comparisons are drawn between results in the two specialities.

CHAPTER EIGHT

The Questionnaire Study 3: Discussion of Results

In Chapter Seven were presented the results of the statistical analyses applied to data from students' and registrars' responses to the questionnaires in endocrinology and neurology. Those results were presented in relation to the statistical analyses performed. In this chapter these results will be discussed in relation to the research hypotheses associated with the questionnaires (Hypotheses 1 to 9, stated in section 5.1 above). Each hypothesis, though stated only once, applies separately to the results for endocrinology and neurology. These, therefore, will each be discussed separately before being considered in relation to one another for each hypothesis. Where necessary or appropriate, the discussion will commence with an explanation of the derivation of or reason for that hypothesis. However, to provide a unifying perspective, we shall preface the discussion of results with a description of the various questions and points about which we wish to make inferences on the basis of the results to be discussed. This will be followed by a brief discussion of the scope of the questionnaires.

8.1 Overview of Issues for Discussion

In the process of discussing and evaluating the results of statistical analyses of data from the questionnaires in endocrinology and neurology, it is hoped to make inferences and reach conclusions about the following points and questions:

- (a) The diagnostic thinking process logically may be divided into two aspects: firstly, the thinking process itself, which we may refer to as the 'cognitive process' or 'cognitive processes'; and secondly, the content of that process in terms of the thinker's knowledge and the outcome of his thinking. This aspect has four identified elements measured by the questionnaires and reflected in the four sections. The relationship between scores on these sections we may call the 'skills structure' of an individual. This term is preferred to the alternative 'knowledge

structure' because it reflects that cognitive skills are measured in the four sections (especially in sections B and D) as well as knowledge itself (especially section A). We therefore hope to be able to infer the nature, relationship and separateness of these two aspects: 'cognitive processes' and 'skills structure'.

- (b) Is the nature of cognitive processes different for endocrinology and neurology?
- (c) Are both the skills structures and cognitive processes of students different in endocrinology and neurology?
- (d) Are skills structures and cognitive processes of students and registrars similar or different in endocrinology and neurology, respectively? The answer to this question may also indicate what are the fundamental pre-requisites for diagnostic acumen.
- (e) Do skills structures alter with clinical practice?
- (f) To what extent do the skills measured in sections A, B and C play a part in the cognitive process of formulating a diagnosis?
- (g) What new conclusions may be drawn about the diagnostic thinking process?

8.2 Introduction to the Discussion

As an introduction to the discussion of results, it may be useful to consider the questionnaires as an instrument for analysis of the diagnostic thinking process, and for identification of differences between students and registrars. Sources of such differences may seem self evident, but we may identify some of them as stemming from amount and type of exposure to clinical phenomena; contextual aspects of knowledge and skill development (learning as a student, as opposed to learning as a clinical practitioner); recency of initial knowledge and skill acquisition; relative rehearsal of knowledge and skills; contextual aspects of knowledge and skill rehearsal (rehearsal in order to become proficient; rehearsal in order to maintain

proficiency; rehearsal as a response to a particular clinical situation, etc.); relative decay of unrehearsed knowledge and skills; and changes in the storage structure, inter-relatedness, use value, meaning or significance and corollaries of knowledge and skills resulting from practice.

With regard to the sections of the questionnaire, it is suggested that the skills tested in sections B and D (interpretation of symptoms and signs, and formulating a diagnosis, respectively) are those most likely to develop with clinical practice. Elstein et al (1978) have shown skill in interpretation of cues to be positively related to diagnostic accuracy, and this latter skill itself is central to clinical practice. It may be argued on logical grounds that the processes tested in section C are subsumed under those tested in section D, thus rendering the differentiation between sections invalid. However, we would argue that the cueing effects of the multiple choice questions renders the process of formulation of diagnostic possibilities and, consequently, testing these, less dependent upon the respondent's own cognition (knowledge retrieval, thinking processes, etc.) than would otherwise be the case. We would argue, therefore, that the cognitive processes pre-requisite for correct response to section C multiple choice questions are of a lower order of cognitive complexity than would be required for successful completion of section D. For this reason, sections B and D are selected for special attention in a number of the research hypotheses.

It might also be useful to give preliminary consideration to the obvious limitations of the questionnaires. The most obvious feature of clinical practice not accounted for by the questionnaires involves the dynamic interactive aspects of clinical problem solving thinking which enable the clinician to guide the flow of information from the patient and requires him to think about information at varying stages of completeness. Section D certainly allows for the generation of successive and changing interpretations of the information given, but does not allow the respondent to take an active part

in testing or following through his interpretations. The order of information flow is fixed. It is likely, then, that section D taps a less complex analogue of real clinical problem solving thinking.

It is also possible that the multiple choice questions omit a study of certain types of knowledge. It is a reasonable contention that changes in storage structure, inter-relatedness, meaning or significance and corollaries of knowledge may accrue from clinical practice. In addition, the knowledge gained from practical, clinical experience, especially from seeing disease manifest itself in different and similar ways in many patients, may well be qualitatively different from the knowledge gained in medical school. If this is the case, we cannot be sure that this has been measured.

Having discussed our preliminary considerations of the possible effects of clinical practice, the relative roles and nature of the four sections of the questionnaires and their limitations, we may proceed to the discussion of results.

8.3 Hypothesis 1 and Hypothesis 5

These two hypotheses are related and so may most effectively be discussed in conjunction. They are stated as follows:

"Accurate interpretation of symptoms and signs is the primary pre-requisite for diagnostic acumen, therefore scores on section B will have more predictive power than scores on section A or C for scores on section D."

"Scores on sections A, B and C combined have no predictive power for scores on section D, either in endocrinology or neurology."

Hypothesis 1 has the implicit assumption that section D tests more nearly than any other section, the results of a diagnostic thinking process, and that sections A, B and C test definable components of that process, which are addressed by undergraduate medical education. All this, of course, is within the bounds and context of the low-fidelity nature of the research instrument, which merely isolates and controls aspects of the process

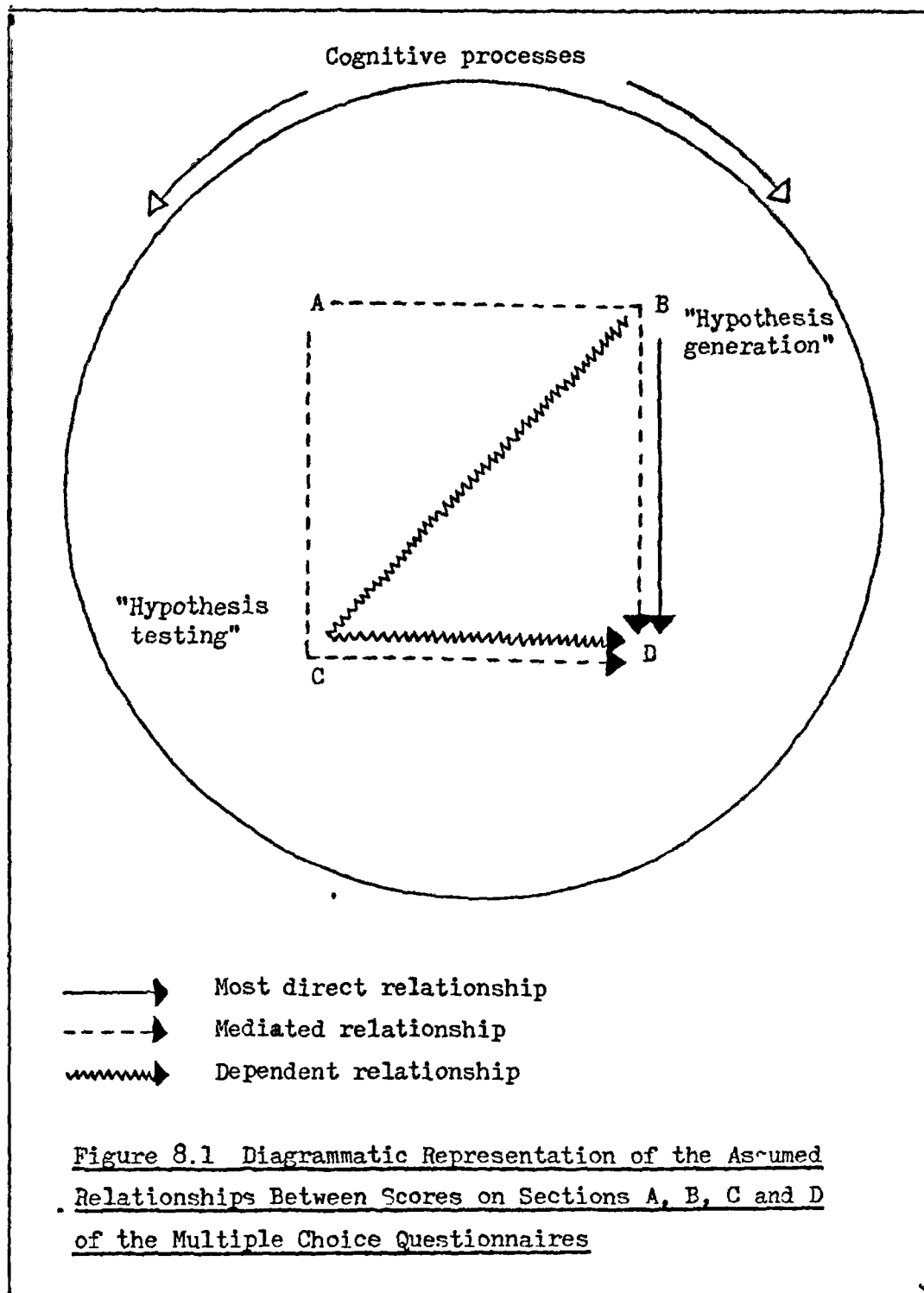
rather than allowing a broad overview of that process in practice, as does the gathering of accounts by stimulated recall.

Section B (interpretation of symptoms and signs) is selected as having most predictive value for scores as section D (formulation of a diagnosis) since the accurate interpretation of clinical information is assumed to be fundamental to the process identified as "hypothesis generation" (see section 3.1) as well as to the ability to test accurately the feasibility of those possibilities. In the absence of accurate interpretation of symptoms and signs, the primary data, the other processes of diagnostic problem solving will be jeopardised. This interpretation is substantiated by the findings of Elstein et al (1978). This principle applies to both students and registrars. Section A (mastery of factual knowledge) is assumed to have a less direct influence on section D, its influence, perhaps, being mediated by the skill of section B. Likewise, scores on section C (selecting and testing diagnostic possibilities) are assumed to be partially dependent upon acumen in the interpretation of clinical information. In addition, the section C skill is assumed to relate closely to the knowledge base tested in section A, without which hypotheses cannot be either formed or tested. Section B, therefore, is selected as the most directly relevant to section D. This set of relationships is represented diagrammatically in Figure 8.1.

It is convenient to discuss Hypothesis 5 at this juncture, since it is closely related to Hypothesis 1 and is answered by the same regression analyses.

8.3.1 Endocrinology

It appears that the skills tested in sections A, B and C, even when combined, play a relatively small part in the cognitive process of formulating a diagnosis. For both students and registrars in endocrinology the combined predictor variables account for relatively little of the variance on



the criterion. For students this value is about 41 per cent (multiple $R^2 = 0.4109$), which is quite low; but for registrars the value is only about four per cent (multiple $R^2 = 0.0395$), which is remarkably low.

It is reasonable to conclude from these low values that one or many relevant variables have been omitted from the set of predictors. In other words, we reflect the statements of 8.2 above, that the process of making a diagnosis, even in its simplified form as measured by section D, has greater breadth, or more aspects, than are measured in the first three sections of the multiple choice questionnaires.

This conclusion must be qualified by our second major finding, that the cognitive processes of students differ from those of registrars. The results of the regression analyses show that registrars rely more heavily on the processes not measured by the questionnaire than do the group of students. Equivalently, we may conclude that the students tend to rely more heavily upon the skills measured by sections A, B and C. It therefore appears to be the case that, whatever the nature of the skill relied upon by registrars, it develops, or is more fully implemented, after the years of undergraduate medical education are over. In fact, we cannot be certain either that students do not possess the same skills or that they possess them, but are, in some way, encouraged to use others. The account gathering study clarifies this issue. The fact of 59 per cent of the variance on the criterion being unaccounted for, however, suggests that final year medical students are already making considerable use of these other skills.

We may conjecture about the nature of the unmeasured variables and suggest that cognitive processes of a dynamic, interpretative nature are applied to the case histories. It would seem likely that as the subject reads each successive sentence and gains each new item of information, his interpretations of the patient's problem may change, as may his ideas about possible or likely diagnoses.. Yet the subject is still only an active, thinking observer, rather than a participant in the clinical encounter. We may consider it likely that the case histories do require the cognitive skills of combining or structuring items of information, of selecting from the array,

or rejecting data, of assessing the relevance of each item and of interpreting its meaning and clinical significance, in relation to some current view of the patient's problem. Or registrars may, possibly, excel over students simply in their capacity to combine these cognitive skills to make complex judgments.

In all our discussions of the possible nature of the unmeasured variables, we are merely reflecting, and taking from, the research findings and suggestions of other workers, as described in Chapters Two and Three above. In particular, we reflect Elstein et al's (1978) model of cue acquisition, hypothesis generation, cue interpretation and hypothesis evaluation. We do not suggest, however, that these form a complete or necessarily accurate set of possibilities. Indeed, we question Elstein's formulation and hypothesise other processes which will be considered further in the study of accounts by stimulated recall. The present study of responses to multiple choice questionnaires forms a baseline against which to interpret those findings.

Our discussions so far, then, has substantiated the broad conclusion that the cognitive processes of students differ from those of registrars. But our discussion has been quantitative and general, in that it is based on the great difference in predictive value of sections A, B and C for section D, between students and registrars. It has only been suggested that the two groups of subjects rely to different extents upon the skills tested and those untested. The McMaster comparative studies of students and physicians (section 3.1.2) do not provide any useful indications of possible explanations for our findings, neither in terms of characteristics of hypotheses (Norman et al, 1977; Neufeld et al, 1976; Barrows, 1976) nor of questions and amount of information elicited (Norman et al, 1977; Rimoldi, 1964; Barrows and Bennett, 1972).

Consideration of Hypothesis 1 may occasion more specific enquiry into the similarities (and therefore differences) in the cognitive processes of students and registrars, and thereby

enable a deeper discussion of the reasoning which culminates in the conclusion that the cognitive processes of students differ from those of registrars. So far, we have shown this to be so in terms of relative contributions of tested and untested variables. Hypothesis 1 states that section B will have the most predictive value for section D, for both students and registrars. This is substantiated but does require some points of qualification. For students, section B accounts for 38 per cent of the variance on the criterion ($R^2 = 0.3797$; Table 7.6) while for registrars this value is only one per cent ($R^2 = 0.0109$; Table 7.8). Although statistically the hypothesis is supported, it would be rash to attach great interpretative value to the registrars' results, since the entire set of predictors accounts for so little of the variance on the criterion and the relative contributions of each predictor are so small and so similar (see Table 7.8).

The second part of Hypothesis 1 states that sections A and C have less predictive value than section B for scores on section D. This is also upheld for both students and registrars, although for the latter group it is again unwise to make any strong inferences. For students, we find that the predictive values of sections A and C, independently of each other and of section B, account for 22 and 13 per cent of the variance on the criterion respectively ($R^2 = 0.2190$ and 0.1353 , respectively). Factual knowledge, therefore, plays a not inconsiderable role in the student's diagnostic thinking processes, while for registrars, this capacity plays a minimal relative role. For students, also, the skill of selecting and testing diagnostic possibilities plays the least part. It would be reasonable to infer from this that students rely on a process of less cognitive complexity than do registrars.

In summary, we may reasonably say that the skills tested in sections A, B and C of the questionnaire in endocrinology play a substantial role in the students' diagnostic thinking processes, although these are heavily over-shadowed by other, untested, processes for registrars. For students, interpreta-

tion of symptoms and signs is the most important skill in terms of formulating a diagnosis. This is followed by mastery of factual knowledge. Such a pattern of skills contributing differentially to the diagnostic process is not at all evident for registrars. This difference in cognitive processes may lead us to consider the role of medical education relative to clinical practice in endocrinology. This question is considered in more depth in Chapter Thirteen.

8.3.2 Neurology

Hypothesis 5, in essence, asks how much predictive power is invested in sections A, B and C for scores on section D. Unlike the endocrinology results, we find that the three predictors account for approximately the same proportion of the variance on the criterion for both students and registrars ($R^2 = 0.2677$ for students; $R^2 = 0.2691$ for registrars). These values are relatively small. In both cases the first variable to be entered into the regression equation is section B, as hypothesised, which also accounts for approximately the same proportion of the variance on the criterion for both students and registrars ($R^2 = 0.2627$ and 0.2341 , respectively). Also in both cases, the subsequent contributions of sections A and C are minimal (increase in R^2 for A and C combined = 0.0050 and 0.0350 for students and registrars, respectively). Considering the independent predictive value of each variable, the only appreciable difference between students and registrars is on section C ($R^2 = 0.1618$ and 0.0222 , respectively).

The independent predictive value of section A is approximately the same for students and registrars ($R^2 = 0.1362$ and 0.1740 , respectively). But, taken overall, the results for students and registrars in neurology are very similar.

The similarities in the final multiple R^2 values do not tell us anything about the nature of the predictors which might account for the variance remaining on section D. We do not know whether the same predictors would be necessary for students and registrars, although the coincidence of the multiple R^2

may suggest that this would be a reasonable hypothesis. Results of the parallel study of accounts by stimulated recall should clarify this issue.

It may be reasonable to conclude from these results that the teaching and learning of neurology is more appropriate to its practice than was the case with endocrinology. However, this conclusion may only be drawn with regard to cognitive processes, and even then only to the extent that the skills of sections A, B and C are encouraged to contribute to the diagnostic thinking process to the same extent as they do in an experienced practitioner. This is not to say that either the manner of their contribution, or combination, or the eventual outcome is similar. Indeed, it is shown below (section 8.6.2) that neurology students are statistically significantly worse at formulating a diagnosis than at any of the skills of section A, B and C. It would therefore be unwise to draw any conclusions about the relative appropriateness of teaching in endocrinology and neurology to the practice of those specialities.

8.3.3. Comparison of Endocrinology and Neurology

The most obvious similarity of results is in the greatest predictive power invested in section B for both groups in both specialities, although for endocrinology registrars this finding is equivocal. In the case of neurology, however, section B has the same predictive power for both students and registrars.

The second point of similarity across endocrinology and neurology is found in the relatively low total predictive power of the variables tested for all groups, and the subsequent conclusion that other skills are necessary and very active in the process of formulating a diagnosis.

In all other aspects, the results for endocrinology and neurology are dissimilar. In endocrinology, registrars rely very heavily on processes other than those measured in sections A, B and C, whereas students rely to a considerable extent ($R^2 = 0.4109$) on those very skills. The potential predictive contribution

of certain variables, independently of all others, also varies between students and registrars in both specialities, but the relationships are different in that endocrinology students differ greatly from their registrars by relying to a much larger extent on their ability to interpret symptoms and signs. In neurology, the major difference is found in the students' relatively greater reliance on their ability to select and test diagnostic possibilities. In both specialities, then, we see differences in the cognitive processes of students and registrars, but these are more pronounced in endocrinology groups, and minimal in neurology differing only on the independent predictive power of section C, which is, in both cases, considerably less important than section B. Our conclusion, then, must be that the cognitive processes of neurology students and registrars, as measured in sections A, B and C, are similar. Of course, this does not necessarily imply either that these groups are similar on the untested cognitive variables, or that they are equally as successful in their use of the variables tested. Indeed, this latter possibility has been refuted, and will be discussed below in relation to the skills structures of students and registrars (section 8.6.2) and to the relationship between medical education and clinical practice (Chapter Thirteen). The different values of R^2 for endocrinology and neurology students will also be discussed in relation to medical education.

The very different values of R^2 for registrars in endocrinology and neurology would seem to imply a speciality specific form of thinking, which may, or may not reflect structural differences in those specialities. It has been argued above (section 3.1.1) that findings of content specificity of thinking are due to the content dependent nature of the variables selected for study. Likewise, it may well be the case that speciality specific cognitive processes reflect speciality specific structures of either information, or the presentation of disease.

In considering Hypotheses 1 and 5, then, we have discovered evidence relevant to points (b), (d) and (f) (section 8.1). Point (b) concerns whether or not the nature of cognitive processes is different for endocrinology and neurology, as evidenced by differences in the stepwise regression analyses of registrar data. In addition, this speciality specific thinking may imply different speciality structures for endocrinology and neurology. Point (d) considers whether cognitive processes, as measured in sections A, B and C, differ substantially between students and registrars in neurology and endocrinology. We find only a great difference in endocrinology, but that this does not imply, and is not accompanied by, a greater diagnostic acumen in neurology students who display approximately the same cognitive processes, as measured in sections A, B and C, as their registrars. Point (f) concerns the extent to which sections A, B and C have predictive power for scores on section D. We find relatively little and conclude that a large area of skill necessary to clinical practice in endocrinology and neurology is not addressed by undergraduate medical education in those specialities.

8.4 Hypothesis 2

"Mastery of the separate skills tested in sections A, B, C and D is presumed to develop differentially during medical education and clinical practice. Therefore, statistically significant interaction effects will be found between groups of subjects and sections of the questionnaire".

This hypothesis is based on the assumption that with clinical practice different dimensions are introduced into any speciality, and the various aspects or skills wax or wane in relative importance. We would therefore expect some skills to increase dramatically while others develop less. In the present study we would not expect to observe any decline in scores on the variables tested since these are all subject or practice related for each individual respondent. The content is therefore of direct relevance to students about to sit the final qualifying examinations and to registrars practising the

speciality in question. Assuming, then, a differential development in skills tested, we could logically expect to find a statistically significant interaction term on a two-way analysis of variance, which would imply that registrars are relatively better than students on one variable rather than another, and that the change in scores is, therefore, not uniform across all skills tested.

8.4.1 Endocrinology

Table 7.13 shows an interaction term which is not statistically significant ($F_1 = 3.0920$; df. 3,272). Our hypothesis is therefore rejected in this instance. Thus, the expected improvement in certain areas based on the different emphasis which clinical practice, of necessity, has in relation to medical education, is not evidenced. We must recall, however, that we are currently considering the skills structures of students and registrars, and not their cognitive processes, which have already been shown to be different. Our conclusion must be that, although registrars score statistically significantly higher than students ($F_r = 88.154$; df. 1,272), they share a similar skill's structure. We have already seen, however, that a similar skills structure does not imply similar use of those skills, similar cognitive processes or similar diagnostic acumen.

8.4.2 Neurology

In this case, hypothesis 2 is upheld. Table 7.14 shows a statistically significant interaction term ($F_1 = 7.7080$; df. 3,272), which suggests that the difference between students and registrars, in variance terms, depends upon the section (or skill) being tested. We can be certain that the statistical significance of this term is due to the effect of very low scores on section D for students (see Table 7.16), since, as for endocrinology, there are no statistically significant differences between any two sections for registrars (see Table 7.25). We must conclude, then, that not only do neurology registrars score statistically significantly higher than stu-

dents ($F_x = 97,6302$; df. 1,272), but they also have a different skills structure. This must be seen in the light of similar use of those skills (that is to say, similar cognitive processes) between students and registrars, as demonstrated by the regression analyses which yielded similar values. This is relevant to point (d) above (section 8.1).

8.4.3. Comparison of Endocrinology and Neurology

The finding of one statistically significant interaction term (neurology) and one term which is not statistically significant (endocrinology) has occasioned a discussion of the relationship between skills structure and cognitive processes in relation to diagnostic acumen. This relationship is best summarised in Table 8.1, although it must be noted that 'different' and 'similar' are relative, not absolute, terms. It would appear that the successful formulation of a diagnosis, as measured in section D, requires both a certain skills structure (in this case, homogeneity at a certain level across the skills tested in sections A, B and C), as well as certain cognitive processes, about which we cannot yet be specific.

Table 8.1 The Relationship Between Students' and Registrars' Skills Structures, Cognitive Processes and Diagnostic Acumen (in relative, not absolute, terms).

	<u>Skills Structure</u>	<u>Cognitive Processes</u>	<u>Diagnostic Acumen</u>
Endocrinology	Similar	Different	Different
Neurology	Different	Similar	Different

Our incidental reference to homogeneity of registrars' scores across all sections (see Tables 7.20 and 7.24) refutes the rationale for Hypothesis 2. It was suggested that some of the skills tested improve dramatically with clinical practice, while others develop less. This is the case for neurology,

where medical education leaves specific skills deficits. It appears that clinical practice requires each of the skills tested to a similar extent, but that the use of these skills, combined into complex cognitive processes, is a separate question and is also a necessary pre-requisite for the formulation of diagnoses. A complementary interpretation of the findings is that, given a necessary minimum threshold of skills and knowledge, diagnostic acumen depends upon factors which have not been tapped by the questionnaires and which, being dependent upon clinical experience, are difficult to teach.

Having rejected the rationale for Hypothesis 2, the same argument cannot be rejected for untested skills, particularly in the case of endocrinology in which the total contribution of those skills tested to the diagnostic process declines markedly with clinical practice, as shown by the results of regression analyses. This is relevant to point (d) above.

8.5 Hypothesis 3

"Given the presumed greater development of skill in sections B and D with clinical practice rather than medical education, observed differences between students' and registrars' scores on these sections will be greater than those observed on sections A and C. However, the greater experience and post-graduate learning of the registrars will cause groups' scores on all sections to be statistically significantly different in favour of registrars".

The first part of this hypothesis is based on the assumptions, defined in section 8.2 above, concerning the relative development of different skills with clinical practice. The second part of the hypothesis is based on the reasonable assumption of continued learning at postgraduate level, and continued rehearsal of learning acquired at undergraduate level.

However, although this rationale is quite acceptable and feasible, it will be shown that the present analyses are not adequate in testing the hypothesis based on it. Gain scores, or the differences shown between students and registrars on

any one section of the questionnaire, are primarily a function of the scores obtained by students. These scores are not constant across sections. Registrars' scores, therefore, can only be interpreted as improvement upon scores achieved at the end of undergraduate medical education. Given the homogeneity of registrars' scores across sections, then, gain scores will be largely a function of medical education, rather than clinical practice. The present hypothesis and analysis is, therefore, partially spurious. Its usefulness is really limited to demonstrating that registrars score statistically significantly more highly than students on all counts. This is relevant to point (e) above.

8.5.1 Endocrinology

The two-way analysis of variance yields a statistically significant value of F_r ($F_r = 88.1540$; df. 1,272) indicating a constant difference between students and registrars across all sections of the questionnaire. Table 7.15 shows that students' and registrars' scores on each section separately are statistically significantly different. The greatest observed differences, however, are on section C (18.26 per cent) and section D (15.34 per cent). Sections B and A show smaller observed differences (10.72 and 7.35 per cent, respectively). Our hypothesis is, thus, partially falsified and partially upheld.

8.5.2 Neurology

Two-way analysis of variance, as for endocrinology data, shows a statistically significant value of F_r ($F_r = 97.6302$; df. 1,272), while Table 7.16 shows that students' and registrars' scores on each section are statistically significantly different, registrars scoring more highly. The greatest observed difference is on section D (24.29 per cent), due to the low mean score of students on this section. The remaining three sections show approximately the same degree of observed difference (10.55, 8.60 and 9.45 per cent for sections A, B and C, respectively). Our hypothesis, again, is partially upheld and partially falsified.

8.5.3 Discussion of Endocrinology and Neurology Results

It has been stated in 8.5 above that the present hypothesis and analyses are partially spurious because of the effect of medical education in yielding disparate scores across sections, while the effect of clinical practice is to equilibrate these scores. Comparison of students' and registrars' scores on each section, therefore, tells us rather more about medical education than clinical practice and, given the homogeneity of registrars' scores, this evidence about the effects of undergraduate medical education would be better gleaned from the direct comparison of students' scores across sections.

Our only conclusion must be that registrars show improvement in all the skills tested to the extent that their scores are statistically significantly higher than those of students for all sections. This may scarcely need proof, but it is worth noting, and is relevant to point (e) above. Again, we may suggest that some necessary minimum threshold of skills and knowledge is reflected.

8.6 Hypotheses 4 and 6 ,

"Sections A, B, C and D test different skills which may be rehearsed in medical education and clinical practice respectively to disproportionate degrees. Therefore, for each group separately, statistically significant differences will be found between scores on the four sections."

"No differences will be apparent between students and registrars in the relationship between scores across sections A, B, C and D either in endocrinology or neurology".

Hypothesis 4 is based on the assumptions defined in section 8.2 above, concerning the relative development of skills with clinical practice and the additional assumption that undergraduate medical education does not make provision for rehearsal of these skills in the same manner or proportions as required by clinical practice. As yet, we make no argument about the appropriateness or not of such a relationship between education and practice, but merely seek indication of the truth or fals-

ity of the hypothesis. Hypothesis 6 is related to this same point and will be answered by the same analyses. The discussion is relevant to point (e) in section 8.1 above, which asks whether skills structures alter with clinical practice.

8.6.1 Endocrinology

One-way analysis of variance has shown that scores across sections are statistically significantly different only for students, and not for registrars (Tables 7.18 and 7.20). This would suggest a differential development and rehearsal during undergraduate medical education of the skills here tested, with a compensation during the years of clinical practice for those less well developed. Our previous findings and discussion would lead us to expect such an observation.

Our hypothesis, then, is only partially substantiated. Results of the Scheffé test for students indicate that the statistically significant value of F is due to the comparison between sections A and C which show the highest and lowest mean scores attained, respectively. This lone statistically significant difference seems to imply a less differentiated approach to the teaching of endocrinology than the teaching of neurology, one result of which appears to be that students in their final year are as adept at actually making a diagnosis in endocrinology as they are at mastering factual knowledge and interpreting symptoms and signs. Even the specific weakness demonstrated in scores on section C (selection and testing of diagnostic possibilities) is only statistically significant in relation to the strength of scores on section A (mastery of factual knowledge), but neither of these is statistically significantly different from scores on sections B or D. The hypothesis for students, then, is substantiated, but weakly so. For registrars it is firmly rejected, since the effect of clinical practice is to equalise, rather than further differentiate between scores across sections.

In dealing with Hypothesis 6, we can state that the relationship between scores across sections is different for students

and registrars, but only in terms of comparing extreme mean scores. Students do show less equality of scores than do registrars but the observed mean values of these differences only reach statistical significance when the highest and lowest scores are compared. The statistically non-significant interaction term reported in 8.4 allows us to accept the relative similarity of skills structures.

8.6.2 Neurology

As for endocrinology, one-way analysis of variance shows that scores across sections are statistically significantly different for students, but not for registrars (Tables 7.22 and 7.24). This again suggests an interpretation of differential development and rehearsal during undergraduate medical education of the skills here tested, with a compensation during the years of clinical practice for those less well developed. Again, Hypothesis 4 is only partially substantiated. However, the neurology results show one very great difference from those in endocrinology, and this is in relation to the derivation of the statistically significant variance ratio for students' results.

The Scheffé test for neurology students reveals that scores on section D (formulating a diagnosis) are statistically significantly lower than scores on any other section. The lack of differentiation weakly indicated in the endocrinology results is therefore strongly denied by the neurology results. The implications of this finding will be fully discussed below (Chapter Thirteen). This conclusion is made more acute when it is recalled that sections A, B and C account for 27 per cent of the variance on section D for neurology registrars and only approximately four per cent for endocrinology registrars.

In considering Hypothesis 6, it has been shown that the relationship between scores across sections is different for students and registrars, due, almost entirely, to students' low scores on section D.

8.6.3 Comparison of Endocrinology and Neurology

The major point of similarity between findings in endocrinology and neurology is found in the statistically significant difference in scores across sections for students, but not for registrars and the concomitant conclusion that clinical practice in each case compensates for any deficiencies left by undergraduate medical education, rendering scores across sections approximately equal.

The specialities differ considerably in the relative strengths and weaknesses of skills displayed by final year students. Endocrinology yields a less differentiated set of scores, while neurology demonstrates a particular weakness, in formulating a diagnosis, which is statistically significantly different from all other skills tested.

Given the clearly articulated philosophy of teaching in neurology, and its conspicuous absence in endocrinology, it seems likely that the differences between students are a function of the differences between teaching strategies. This is further discussed below (Chapter Thirteen). However, the possibility that these differences are actually due to some fundamental contrasts between the specialities themselves cannot be totally discounted.

8.7 Hypothesis 7

"No major differences will be observed between results for endocrinology and neurology".

This hypothesis has been answered in the third part of each of the above sections. These findings may be summarised as follows.

Firstly, the homogeneity of registrars' scores suggests that a major difference between the specialities considered is in the associated skills structure of students and so is of pedagogic origin.

Secondly, for registrars, sections A, B and C have much greater

predictive power for scores on section D in neurology than in endocrinology (total $R^2 = 0.2691$ and 0.0395 , respectively). In addition, for endocrinology very different values of R^2 are observed for students and registrars (0.4109 and 0.0395 , respectively), while for neurology these values are similar (0.2677 and 0.2691 , respectively). Different values of R^2 suggest different degrees of reliance on the skills tested in formulating a diagnosis. The greater the R^2 value, the greater the reliance.

Thirdly, in making a diagnosis, in endocrinology students differ greatly from registrars in their reliance on their ability to interpret symptoms and signs ($R^2 = 0.3797$ and 0.0109 , respectively), while neurology students differ greatly from registrars in their reliance on their ability to select and test diagnostic possibilities ($R^2 = 0.1618$ and 0.0222 , respectively). In both cases, the students' reliance is greater. This difference would seem to reflect differences in teaching strategy.

Fourthly, in endocrinology, the teaching strategy appears less differentiated, or perhaps less analytical of component skills or processes, than in neurology. This is reflected in the students' relative scores across sections.

Fifthly, in endocrinology the total contribution of the skills tested in sections A, B and C declines markedly with clinical practice. In neurology, that total contribution remains constant, although its component parts may alter in role or importance.

8.8 Hypothesis 8

"No developmental pattern will be apparent, such that differences in scores between students and registrars will not be interpretable in terms of differences between medical education and clinical practice".

The existence of some developmental pattern may be inferred from a comparison of results of students and of registrars. It has already been suggested that the relative scores of students across sections are primarily a function of medical education.

So far as any developmental pattern is apparent, we may really only make one point on the basis of this part of the study. It is intended that the results of the accounts gathered from stimulated recall will furnish us with more direct evidence. Our only possible conclusion from the questionnaires, however, must be that, whatever the process of development is, it is distinguished by one characteristic, which is the tendency to equalise skill in each of the sections tested, regardless of the starting point of each. In endocrinology, there is the added characteristic of an increasing development of, reliance upon or use of skills not measured by the sections of the questionnaire. We thus interpret the differences in scores between students and registrars in terms of differences between undergraduate medical education and clinical practice. We contend that different teaching strategies in endocrinology and neurology give rise to different patterns of scores but that, with clinical practice, these different patterns evolve into similar ones. Let us consider the effects of medical education a little more deeply. Our discussion will be relevant to conclusion (d) below (section 8.10), which states that the process of medical education yields different patterns of characteristics (skills structures and cognitive processes) in its recipients than does the process of clinical practice in its executors, and that this unequal relationship between medical education and clinical practice manifests itself differently in endocrinology and neurology.

Endocrinology as a speciality has no clearly articulated philosophy of or approach towards teaching which can be found either in the literature or by observation of its practice. No clearly similar teaching strategy would predictably be observed in all teachers of endocrinology. No one skill or set of skills seems to be singled out for particular attention. This we may term an undifferentiated, unanalytical or holistic approach to teaching, and we have found that this results in a similarity of scores across sections (only the comparison of highest and lowest yielding any statistically significant difference). In turn, this results in a constant improvement across all

sections with clinical practice and a non-statistically significant interaction term. We therefore postulate similar skills structures in endocrinology students and registrars. However, we have shown that similar skills structures are not necessarily accompanied by similar cognitive processes (see section 8.4.3 above). This is particularly the case in endocrinology, where students rely on the skills tested to a far greater extent than do registrars.

Turning to neurology, this speciality has a clearly articulated philosophy of teaching (see 5.3.1 above) which is differentiated in its approach, paying attention to particular aspects of the speciality. In turn, this approach produces greater differentiation of scores across sections. In particular, scores on section D (formulation of a diagnosis) are statistically significantly lower than scores on all other sections. Relating this finding to the structure of neurology teaching, we may consider whether or not it is reasonable to expect diagnostic acumen to develop in parallel with learning to:

- 1) localise the lesion;
- 2) define the general pathology;
- 3) define the special pathology;
- 4) make as accurate a diagnosis as possible on the clinical evidence; and
- 5) consider carefully the investigations required. (See 5.3.1).

The crucial point is surely the fourth one. This teaching strategy seems likely to make the student consciously aware of the tools of his trade as a diagnostician, and of the logical relationship of these tools. However, an awareness and understanding of the tools of the trade is not a sufficient condition for the practice of a craftsman, who is defined by his skillful manipulation of those tools in shaping the final form of his work. The statistically significantly lower scores on section D for students may imply a need for such skill in manipulating and using those elements of his trade so identified for him.

Undergraduate medical education in neurology, then, yields a skills structure in students unlike that of registrars. However,

dissimilar skills structures are not necessarily accompanied by dissimilar cognitive processes. Indeed, in neurology, the cognitive processes of students and registrars are shown to be similar.

It is clear, then, that, as conclusion (d) below (section 8.10) states, the unequal relationship between medical education and clinical practice manifests itself differently in endocrinology and neurology. In the former only the skills structures of students and registrars are similar, whereas, in the latter, only the cognitive processes are similar. Thus the different teaching strategies yield different skills structures and different cognitive processes in students. It is also clear that clinical practice in each speciality requires similar skills structures but different cognitive processes and that whatever is deficient in the final year student is compensated for by clinical practice. It would be inappropriate to consider whether one form of teaching strategy appears a better preparation for clinical practice than does the other. We may only make three comments. The first is that they are different and yield different results. The second is that neither directly addresses itself to the area indicated by the large proportion of variance on section D left unaccounted for by scores on sections A, B and C. In addition, education in endocrinology has a greater distorting effect on this value than does education in neurology. It may well be that the more differentiated and analytical approach of the latter at least enables students to begin to develop and use cognitive processes similar to those evidenced by the practitioner, even though the results of those similar cognitive processes are not as good as those of the registrars. This may be due to the processes themselves or to lesser knowledge and skills (that is, lower scoring on sections A, B and C), since cognitive processes require some content on which to operate. Meanwhile, in endocrinology, it is possible that the students initially can identify only the content aspects of the holistic approach to teaching and learning and rely on these rather than appreciating the structural aspects of the speciality and thereby

relying more heavily on their own cognitive operations on the given content. This would seem a plausible explanation of the very different predictive values of sections A, B and C for endocrinology students and registrars (multiple $R^2 = 0.4109$ and 0.0395 , respectively).

Finally, despite the apparent deficiencies of undergraduate education in the two specialities, it seems that, at least for some, that education provides a suitable basis on which to build improvement and amendment as necessary when clinical practice ensues.

8.9 Hypothesis 9

"Results will have no interpretative value for current descriptions of the diagnostic thinking process".

Although we cannot expect that such a method as multiple choice questionnaires will yield any detailed or specific indications of the nature of the diagnostic thinking process, it does furnish us with some broad indices.

Firstly, and most importantly, our results indicate the extent to which the diagnostic thinking process is dependent upon skills other than those of mastering factual knowledge, interpreting symptoms and signs and being able to select and test diagnostic possibilities. That this is the case has long been the reasonable contention of many workers, but the great extent of the use of these other skills has not before been shown.

Our present results demonstrate most clearly that the skill of cue interpretation is primary in predicting skill in formulating a diagnosis for all groups of subjects, although the value of that variable in predicting endocrinology registrars' scores is remarkably low. This latter finding is somewhat at variance with the findings of Elstein et al (1978) whose own results suggest that diagnostic accuracy in clinicians is related to both thoroughness of cue acquisition as well as accuracy of cue interpretation, despite these two variables being uncorrelated. They present no results for students. Both cue

acquisition and cue interpretation are seen by Elstein et al (1978) as acquired processes stored in memory until needed, with cue acquisition depending heavily on "routinised knowledge of history taking and physical examination, which permits a physician to select smoothly from a battery of questions and manoeuvres stored in memory". Our main interest is in cue interpretation, since the case histories' provision of information ensures thoroughness of cue acquisition. We concentrate, therefore, on the different findings concerning cue interpretation. A number of explanations for this apparent difference in findings may be forwarded.

Firstly, Elstein et al's (1978) results are based on high-fidelity patient simulations, whereas ours are from low-fidelity multiple choice questions and written case histories. There is, therefore, the possibility of incomparability. Secondly, Elstein et al's subjects were required to think aloud and give episodic reviews during the clinical interview. This may well have spuriously elevated scores on either of the independent variables under consideration. The researchers do, themselves, mention the possibility of unreliability of measurement.

The most feasible explanation, however, is that the difference found between our results and those of Elstein et al (1978) is due to different statistical analyses. The Michigan conclusions are based on univariate analysis of variance, comparing diagnostically accurate and inaccurate groups on a number of variables separately. The only two variables which were found to show statistically significant differences were cue acquisition and cue interpretation, the accurate group scoring more highly in each case. We also find this result (see Table 7.15). We therefore question Elstein et al's (1978) interpretations of their results. Univariate analysis of variance can tell us about differences between groups on one variable at a time. But to extrapolate from there to suggest that there is a causal relationship between the characteristic which defines the groups and a number of variables on which those groups are found to be statistically

significantly different is fallacious and inadmissible. The most that Elstein et al's (1978) data and analysis admits of is recognition of the demonstration that groups which differ in diagnostic accuracy, also differ in cue acquisition and cue interpretation. Our own results are in accord with this finding. We therefore accept the results of our regression analyses, and note that the students appear to place considerable reliance on their ability to interpret symptoms and signs, whereas registrars appear to rely on other skills, not here measured. Our findings also suggest that the cognitive skill of formulating a diagnosis is rather greater than the two important elements identified by Elstein et al (1978).

With regard to the contentious area of content specificity of thinking, our current results give relevant information. Section 3.1.1 indicated that there are opposing views about this aspect of the diagnostic thinking process. Elstein et al (1978) find indication of case related thinking, or, at least, case related effectiveness (which may seem hardly surprising), while Berner and Bligh (1974) seriously question the concept of and evidence for content specificity. Our own interpretation of the conflicting viewpoints suggests that some processes measured are content dependent (for example, the nature and number of hypotheses) while other processes are not (for example, the psychological process of actually generating hypotheses) and, depending upon to which of these processes the researcher addresses himself, he will elicit results that appear to support or refute allegations of content specificity of thinking processes. In some ways, therefore, the question when put at the level of cases, is a spurious one. It seems more reasonable to consider the matter at the level of specialities, so long as it is permissible to argue that there actually are differences between the content, structure and presentation of information as well as in matters of investigation.

Given the apparent large effects of teaching and learning, we may only refer to the results for registrars in order to gain indications of content specificity of thinking. We have

already noted (section 8.3.3 above) that the very different values of R^2 for registrars in endocrinology and neurology would seem to imply some important differences in speciality structure and, more importantly, in ways of thinking about those specialities. These results only enable statements to be made about the relative contribution of the skills tested to the overall process and do not provide any indication of how they are used or what other untested processes are involved. It is possible to conclude, however, that the gross difference in the predictive values of the combined variables (four per cent for endocrinology registrars; 27 per cent for neurology registrars) gives clear indication of fundamental differences in diagnostic thinking processes.

What novel conclusions may we draw, then, about the nature of the diagnostic thinking process? Firstly, we have identified two aspects of that process, here called "skills structure" and "cognitive processes", and we have been able to define some relationship between them, although that definition is, as yet, still incomplete and imprecise. Secondly, we have been able to demonstrate the extent of the influence and role of the skills structure and cognitive processes respectively. Thirdly, we have been able to demonstrate the necessity for diagnostic acumen of both an adequate skills structure and appropriate cognitive processes. Fourthly, we have shown that both are amenable to shaping by educational processes as well as by clinical experience. Finally, we have shown that cognitive processes, but not skills structures, are, in these instances, speciality specific, a formulation preferred to that of case specificity which is rejected.

8.10 Summary and Conclusions

The discussion of results has been structured around the nine relevant research hypotheses. We may now summarise that discussion by drawing it together in terms of more broad conclusions.

- (a) The results of regression analyses provide information concerning the subjects' thinking processes in formulating

a diagnosis. These are here referred to as "cognitive processes". The results of analyses of variance and associated tests provide information concerning the subjects' levels of knowledge and cognitive skills. These latter results, which concern absolute scores, variances and the relationship between these across sections, are here referred to as "skills structures".

- (b) The nature of cognitive processes in formulating a diagnosis seems to be different for endocrinology and neurology, as evidenced by the results of regression analyses on registrar data. This may also imply differences between these specialities in the structure of content matter (here, referred to as "speciality structure").
- (c) The results of regression analyses and analyses of variance are different for students in endocrinology and neurology. It may be inferred that medical education as delivered in each speciality has a determining effect for the structure of knowledge and cognitive skills (the skills structure) as measured in sections A, B, C and D and shown by the inter-relationship of scores on these sections. In addition, the manner in which students use or combine these skills in formulating a diagnosis differs according to the structure of education delivered.
- (d) In neurology, students and registrars evidence relatively similar cognitive processes, but relatively different skills structures. In endocrinology, students and registrars are relatively similar in skills structures but relatively different in cognitive processes. In both endocrinology and neurology, registrars' skills structures demonstrate the property of homogeneity. In both endocrinology and neurology, students' and registrars' scores are statistically significantly different on section D (formulation of a diagnosis) in favour of registrars. In both endocrinology and neurology, registrars' scores are statistically significantly higher than those of students on sections A, B and C also. Thus, the fundamental pre-

requisites for diagnostic acumen include homogeneity of skills, appropriate levels of skills and appropriate cognitive processes. In addition, it seems that the process of medical education yields different patterns of characteristics in its recipients while the process of clinical practice tends to make its executors more homogeneous; and this unequal relationship between medical education and clinical practice manifests itself differently in endocrinology and neurology.

- (e) Although the skills structures of students at the end of their period of undergraduate training shows statistically significant differences in scores across sections, clinical practice equilibrates the absolute levels (and variances) of the skills measured in sections A, B, C and D, in such a way that registrars show no statistically significant differences between scores across sections in either endocrinology or neurology, while scoring statistically significantly higher than students in all sections. This may indicate a necessary threshold of skills and knowledge.
- (f) For students and registrars in endocrinology and neurology, the skills tested in sections A, B and C, even when combined, play a relatively small part in the cognitive process of formulating a diagnosis. This is most marked for endocrinology registrars. It is concluded that undergraduate medical education in the two specialities considered fails to address itself to a large area of skill and knowledge necessary in the clinical practice of those specialities. This has not previously been quantified.
- (g) A number of novel conclusions may be drawn about the nature of the diagnostic thinking process. Firstly, we have identified two aspects of that process, here called "skills structure" and "cognitive processes", and we have been able to define some relationship between these two elements which together form the "cognitive structure". Secondly, the extent of the influence and role of the skills structure and cognitive processes is demonstrated. Thirdly, is

demonstrated the necessity for diagnostic acumen of both an adequate skills structure and appropriate cognitive processes. Fourthly, is shown that both are amenable to shaping by educational processes as well as by clinical experience. Finally, we have shown that cognitive processes but not skills structures, are for endocrinology and neurology, speciality specific, a formulation preferred to that of case specificity which is rejected.

The implications of these conclusions for medical education will be discussed in Chapter Thirteen.

CHAPTER NINE

The Account Gathering Study 1: Description, Validity, Reliability and Data Analysis

This chapter presents a detailed description of the method of account gathering by videotape stimulated recall of a clinical interview. Comparisons with the McMaster and Michigan uses of the method will be made. This is followed by a discussion of issues concerning the validity and reliability of the method and analysis of the data.

9.1 Description and Discussion of the Implementation of the Method

The method is, clearly, closely related to those of the Michigan and, more particularly, the McMaster groups as described in Chapter Three above. However, the present study did not require subjects to think aloud as the Michigan studies did, neither were the McMaster complexities of observers' interpretations and analysis of interpersonal behaviours included, the latter not being the domain of our study. The probable jeopardising effect of thinking aloud has been discussed above (section 3.1.1) and was omitted for the reasons given. Observers' subjective interpretations were considered unhelpful, in that a lack of congruence between observers' and subjects' reports of the diagnostic process would largely be uninterpretable, or, at best, would present severe judgmental problems.

The present study has also omitted many of the quantitative parameters applied by the Michigan and McMaster groups, or, where these are included, has used them only as an approximate guide. It is considered that oral accounts from untrained subjects do not represent data amenable to detailed quantification or quantitative analysis. This matter of potentially spurious quantification has been discussed earlier (see 3.1.1). Our use of the method is primarily to give a qualitative, not quantitative account of the diagnostic thinking process, complementing the quantitative results of the questionnaire study.

Our data is therefore primarily derived from the actual accounts given by subjects in response to the videotape stimulus. The methodology used to elicit these accounts is more standardised and rigorous than the methodologies applied in the Michigan and McMaster studies.

The preparation for each stimulated recall, account gathering session began with the selection of a suitable patient (see 9.2.3. below). Our use of real patients rather than simulators represents a fundamental difference from the Michigan and McMaster approaches. The patient was advised that he/she would be interviewed by a doctor and that the interview would be videorecorded, but the recording would only be viewed by the doctor and one other person. The subject was invited to take part in a project concerned with the way in which doctors make diagnoses. Each subject was told that the session would involve interviewing a patient, but not performing a physical examination, and then watching and discussing a videorecording of that interview. Instructions were given to the subject to take a normal history. No time limit was given, but if the subject pressed for some indication, then 20 minutes was given as an average length of time, with the promise that he could take as little or as much time as would be normal. In the cases of both patients and subjects, the opportunity was given to decline involvement in the study. No person took up this opportunity.

The interview took place in a side room. The camera was in evidence, but the videotape recorder was in an adjacent room with a monitor screen for simultaneous viewing by the experimenter. The subject and patient, therefore, were alone in the interview room, while the experimenter viewed the encounter simultaneously, noting the content of the interview as it progressed, in terms of the questions asked and information either elicited or given spontaneously. These notes were used to guide the experimenter during the stimulated recall

sessions and as part of the final transcript.

An example of the experimenter's notes taken during the simultaneous viewing of one encounter between a patient and a house officer is given in Appendix 17. The basic information content of the interview, and all questions asked by the subject are recorded. These notes are not intended as a complete summary of the interview, although often they could have been used as such. Their main purpose was to assist the process of stimulated recall by allowing a reasoned assessment of where the videotape should be stopped to allow discussion, instead of relying entirely on stopping at fixed intervals.

Upon completion of the subject-patient interview, the patient returned to the ward, the videotape was rewound, and the subject came into the adjacent room to view the recording with the experimenter. No other person was present. At this stage, the subject was given the following information by the experimenter:

"I'm going to replay for you the videorecording of that interview. I'll stop the tape at intervals, and when I do, I'd just like you to say why you were asking the questions you asked in that section, what line of enquiry you were following, what you were thinking about the patient's problem at the time and so on. You might find that I stop the tape in all the wrong places for you. If I do, then just tell me, or if you want to stop it yourself to say something, then just say so".

Having ascertained that the subject understood these instructions, an audiotape recorder was switched on and the videotape playback commenced.

The identification of points at which the videotape was to be stopped to gather the subject's account of his thinking processes was based on the application of a number of criteria. Experience of practice runs had shown that to stop the tape according to either set time intervals or the number of exchanges between subject and patient resulted in contextual

difficulties for the subject, often interrupting the flow of a series of related questions, or failure to stop when one group of questions had obviously ended. Likewise, the length of the replay was best determined in relation to its content. While frequent stops may be necessary when the subject is enquiring into the history of the present complaint, a routine family or social history which yielded nothing pertinent whatsoever would not require many stops. It often occurred that, as the interview progressed from specific to routine enquiries, the length of replay also increased. However, a delicate balance had to be maintained throughout the session between holding the subject's interest and motivation for giving the account by allowing frequent occasions for comment, and inducing a set for unthinking responses by causing the subject to have to reply 'just routine, nothing important' too often. Stopping the tape, therefore, was very much a matter of the experimenter's judgment of a balanced application of the following non-quantifiable criteria:

1. Replay periods should not be so short, that the context of the content is obscured.
2. Each replay period should have an information load appropriate to the nature of that information; as information is progressively more non-contributory (as with routine questioning which yields routine answers), the length of the replay may increase.
3. The replay period should not be so long that the subject becomes a passive viewer of the process rather than its active interpreter.

Tables 9.1, 9.2, 9.3. and 9.4. show comparisons between the three groups of subjects on the length of patient interviews, length of replays, number of stops made during the account gathering sessions, and total number of utterances from subject and patient made during each replay period. An utterance is defined as a statement or question made by either patient

or subject, apart from those serving no purpose other than the specific communication of a request for information. This excludes such utterances as "I see", "Uh-huh", etc.

Table 9.1 Accounts by Stimulated Recall. Length
(in Seconds) of Patient Interviews for Students,
House Officers and Registrars

<u>Group</u>	<u>Number</u>	<u>Mean</u>	<u>S D</u>	<u>Range</u>
Students	22	1098.2	321.6	395 - 1566
House Officers	22	974.3	383.2	347 - 1880
Registrars	22	1076.5	577.2	384 - 1888

Table 9.2 Accounts by Stimulated Recall. Length
(in Seconds) of Replay Periods for Students,
House Officers and Registrars

<u>Group</u>	<u>Number</u>	<u>Mean</u>	<u>S D</u>	<u>Range</u>
Students	22	73.7	38	15 - 219
House Officers	22	86.7	49.7	8 - 285
Registrars	22	73.9	45	6 - 247

Table 9.3 Accounts by Stimulated Recall. Number
of Stops Made for Account Gathering for Students,
House Officers and Registrars

<u>Group</u>	<u>Number</u>	<u>Mean</u>	<u>S D</u>	<u>Range</u>
Students	22	14.4	4.2	6 - 26
House Officers	22	11.1	5	8 - 29
Registrars	22	12.4	8.1	6 - 38

Table 9.4 Accounts by Stimulated Recall. Total
Number of Utterances Made During Each Replay Period
for Students, House Officers and Registrars

<u>Group</u>	<u>Number</u>	<u>Mean</u>	<u>S D</u>	<u>Range</u>
Students	22	21.4	9.2	2 - 36
House Officers	22	20.4	10.6	2 - 38
Registrars	22	15.3	9.4	3 - 38

The videotape playback having commenced, the first stop invariably was made after the first couple of exchanges or less, when the patient had given his first response to the subject's enquiry: "What brought you in to hospital?" or "What's the trouble, then?" or some similar question, designed to elicit the patient's main complaint. Having stopped the tape for the first time, the experimenter then said:

"I won't normally stop the tape this often, but that was the beginning of the interview and the first piece of information you really got. Do you usually begin your interviews in that way?"

The purpose of this remark was merely to establish rapport between experimenter and subject and to ensure that the subject would direct his comments along the required lines. Especially, it was often found necessary to reassure the subject that his interpersonal handling of the patient, and the doctor-patient relationship, were not of any interest to the experimenter. Having established these points, the subject was then asked what he had thought on being given the first piece of information by the patient.

The account gathering session then continued with regular replay periods and stops. Wherever necessary or helpful, the experimenter reminded the subject of the content of the replay period just viewed or stimulated the account with non-direct-

ive questions or comments. The experimenter's interventions were of the order: "What about his shortness of breath?", "What were you thinking about here?", "So you got 'No' to cough, 'No' to sputum, 'No' to bringing up blood, and 'Yes' to smoking". The entire process was as non-directive as possible, merely prompting and encouraging. The only deviation from this pattern involved either a brief review of what the subject had just said, or, occasionally, an enquiry about how the subject's current stage of thinking related to previous thoughts about the patient. These enquiries were of the order: "What about your previous ideas of ?", "How's all this fitting in with your previous picture?".

At the end of the account gathering session, the experimenter asked three final questions:

1. "Please could you give a review or summary of this patient?"
2. "If you could have examined this patient and ordered investigations and tests, what would you have done; and what might you have expected to have found?"
3. "How typical of you was that interview in its general structure?"

After each session, the audio-cassette was transcribed with the experimenter's notes inserted at the appropriate junctures. The audio-cassettes were retained for the stage of data analysis, since they also provided an audio-recording of the videotape replay periods themselves. For examples of final transcripts with the experimenter's notes from account gathering sessions with a student, a house officer and a registrar, see Appendices 18, 19 and 20 respectively.

9.2 Validity of the Method

Problems for the establishment of validity and reliability of the method of account gathering using videotape stimulated recall, are related to both practical and theoretical questions,

some of which have already been discussed. A primary question relates to the completeness and quality of the accounts given. Do the accounts reflect an actual thinking process or a mediated, rationalised interpretation of a thinking process? How complete are the accounts? To what extent may the experimenter quantify the accounts given? To what extent is there a need, or justification, for inference of further, unreported, underlying cognitive processes? Other questions are related to the mechanics and design of the method itself. To what extent are the patients comparable? Will incomparability, and the uncontrolled nature of their contributions, jeopardise the research design? To what extent may videorecording interfere with behaviour? Finally, we must consider in what ways it is actually possible to establish validity and reliability of the method and the data. It is to such questions and problems as these that the following discussion is addressed.

9.2.1 The Method of Introspection

In order to set the current research method in its appropriate context, it may be useful to consider what type of process we are initiating in eliciting accounts or verbal report protocols by a method of stimulated recall. The present method is unquestionably related to the classical forms of introspection which are typified by the work of Wundt during the latter part of the nineteenth century (see Miller, 1962). Introspection has been defined as: "Observation by an individual of his own mental processes; systematic self observation; as employed in psychological experiment, it is most frequently immediate retrospection, rather than introspection in any strict sense" (Drever, 1952). The present method differs from this slightly in that the retrospection is assisted and guided by agents other than the subject himself and is not immediate in the sense of providing an account of a small portion of the thought process or of psychological responses to stimuli as they have just occurred. The method of stimulated recall allows the entire problem solving or thinking process to occur without

interference before the subject gives an account of any part of it. Despite such differences, however, the method may be described as introspective and its theoretical validity may be evaluated, in part, with reference to general discussions of introspection as a generic method.

Radford (1974) provides a wide discussion of the role of introspection in psychology. He distinguishes between 'introspections as data' and 'introspection as a method'. Considering the former, he suggests that verbal reports can be recorded in the same way as we would record lever pressing or test scores, but that the data must be subject to such checks as may usually be applied to experimental data and must yield satisfactory levels of significance. In addition, he suggests that the data be not taken at its face value, but rather should have its relationship to some other variable established empirically. These two conditions make certain assumptions about the nature of introspective data, which are difficult, if not impossible, to substantiate. It appears to be assumed that the data is inherently quantifiable, and that, if this were so, quantification would be a reasonable and meaningful course of action. Neither assumption is universally true. Introspective data is primarily qualitative, eminently suited to process tracing research. Although this does not exclude the possibility of quantitative measures, neither does it always admit of forms amenable to statistical analysis. In addition, it is indisputably the case that the researcher cannot know what proportion of thought is being reported. Quantification is, therefore, not necessarily a meaningful course of action in all cases. We would suggest that introspective data should primarily be treated as qualitative and, where quantification is possible, interpretation of results should be conservative. Lack of adherence to these tenets is seen as a severely limiting factor in the Michigan and McMaster research. The present study, however, applies these principles throughout.

Radford's (1974) second category has been identified as 'introspection as a method' and he defines three sub-groups within this category. Firstly, there is self observation, in which the subject aims to observe his own mental events as a behaviourist observes responses. Secondly, there are self reports, in which the subject tells of his experiences without trying to be objective. Thirdly, there is thinking aloud. The present method seems not to be adequately described by any of these three categories, but is like Peel's (1971) method of assessing judgments by the subject's explanation of them. The current method elicits the subject's explanation of his own observed behaviour, in terms of its determinant, preceding and subsequent cognitive operations. Such explanation is unlike the verbalised self awareness of Radford's first and second categories. The present method, in addition, has no elements of any thinking aloud technique which has been discussed and criticised above (section 3.1.1) in relation to the Michigan studies. It is interesting to note in passing, however, de Groot's (1965) report that Duncker (1945) did not consider thinking aloud to be an introspective technique at all, but merely a verbalisation of mental activity.

Despite Radford's consideration of methods of validating 'introspections as data', he does not recommend any similar processes for 'introspection as a method'. From the point of view of research design, however, validation of the method is surely as important an issue as validation of the data. The current study, therefore, has paid considerable attention to standardisation of procedures on the basis of a theoretical rationale and validation of the method of account gathering as put into practice. Validation procedures are, of necessity given the nature of the method and the data, less rigorous than those described for the questionnaires in endocrinology and neurology (see Chapter Six), but this does not imply any diminished degree of adequacy or credibility of the data. The greatest potential sources of invalidity are inappropriate

manipulation or interpretation of data, and unstandardised methodology. We have assiduously attempted to elude both these pitfalls. As Radford (1974) concludes, introspective techniques yield data about experience which would otherwise be inaccessible. It may illuminate facts that might otherwise be overlooked, or it may stimulate new questions. His borrowed aphorism that "je prends mon bien ou je le trouve" is surely too casual an approach to the use of a positive and positively productive research method.

Before reviewing other uses of introspective techniques related to the present study, one major criticism and four less serious potential defects of the method must be discussed. Initially, let us consider the four potential defects discussed by Osgood (1953) in his classical text on methodology. Firstly, he maintains that data from introspection are "patently unverifiable". This criticism, however, is less strong than it appears, particularly in cases of repeated, independent reports of responses to similar stimuli, or of repeated responses of the same subject to a controlled stimulus. In other words, veracity of introspective data may be established in some circumstances in the same way as with other kinds of data. For cases where only one subject gives a response, the data are unverifiable, but this is not the case in the present study which includes large numbers of subjects (in relation to the more usual small scale of introspective studies). In addition, the content area is strictly circumscribed for the subject, and has the external concomitant of an objective recording of the subject's behaviour of which his introspections provide an explanation. Theoretical consideration of the nature of the problem solving process also allows an estimation of the veracity of subjects' reports.'

Osgood's second criticism is that many relevant data are unavailable to the method, yet this criticism can be levelled at any research technique and requires only that the experimenter be aware of the limitations of his method, since

reliability and validity of research method does not attach to the method per se, but also to its use, implementation, application and the manner of interpretation of ensuing data. Our dual research methodology, standardised technique and conservative interpretation of accounts, is based on awareness of such possible limitations of the method of introspection, even when used in conjunction with a stimulus to memory and reconstruction of mental events.

Osgood's third point is that language is not a mirror of thought, and is only as fine a tool as the discriminations it contains. Although this criticism surely applies to the verbal description of emotions, feelings or sensations, it would seem less likely to be a jeopardising factor in such a subject matter as making a medical diagnosis. Finally, Osgood suggests that only the effects of thought, not the process itself, can be observed. But this question of 'observing what is doing the observing' is largely presented as a philosophical one. It will be shown to be clearly the case in this study that subjects can explain their own problem solving thinking processes to a degree unavailable by mere observation of their problem solving behaviour. For our purposes, the effects of thought are seen on the videotape recording, while clues to the underlying process are given in the accounts prompted by stimulated recall.

Related to Osgood's four points is a major criticism of the method of introspection put forward by Evans and Wason (1976) on the basis of a dual process model of thinking. One prediction of the model is that subjects' introspective justifications of their reasoning behaviour (Type 2 process) do not reflect insight into the underlying thought processes (Type 1 processes), but are essentially rationalisations. It is postulated that subjects are unaware of Type 1 processes, while the conscious Type 2 process is obtainable through introspection. Thus a Type 2 process may rationalise the behavioural outcome of a Type 1 process (Evans, 1976). Fellows (1976)

rejects this viewpoint on two grounds; firstly, if the subject's reports are to be dismissed on such grounds, then logically the experimenter must dismiss his own explanations in the same way; and, secondly, a subject cannot be taken as 'hapless' or 'a fool' while the experimenter is seen as 'all wise'. However, the apparently opposing viewpoints are not irreconcilable in all circumstances. Certainly, where there is no corollary evidence other than a subject's own self reports, it cannot be established that those reports are not merely rationalisations. But where external concomitants are available, and the content of reports is guided and circumscribed, as in the present study, some proper estimation of the worth of the data is possible.

The above controversy is in relation to the 'thinking aloud' technique about which we have already expressed serious reservations. In essence, the Evans' (1976) viewpoint is that the processes underlying problem solving behaviour can only be discovered by an experimental inferential procedure. We would agree with that viewpoint. In addition, Radford's (1974) evaluation of introspective data as being of the same order as lever pressing or test scores, to be subject to checks and validation procedures, and our own standardised approach and large numbers of subjects, would certainly permit use of those data as a basis for inferences about unreported processes. They also have validity in their own right. It must be emphasised, however, that this conclusion cannot automatically be extended to introspective studies in other subject areas or to those which lack the elements of circumscription and stimulated recall here present.

The method of introspection has been widely used in studies of judgment in clinical psychology (see Goldberg, 1968) and in other psychological research. Newell and Simon (1972) describe the analysis of verbal report protocols for theory verification as "a sort of hallmark of the information processing approach". Verbal behaviour as data is particularly valued

for its high output rate. Elstein et al (1978), who, like the McMaster group, use introspective data themselves, point out that Piaget's clinical method has rendered verbal reports "legitimate data", and, reflecting Evans (1976) and Evans and Wason (1976) refer to the data as reporting "the contents of consciousness". De Groot (1965) provides an historical overview of the uses of introspection, while Valentine (1978) also provides an account of its historical development and the use of the method by different schools of psychological thought. In addition, he discusses the theoretical problems, reviewed above, of the infallibility or not of introspections and the practical problems of inaccessibility of unconscious thinking processes and mnemonic distortion of reports. Although Valentine's discussion is largely related to thinking aloud techniques, his conclusions reflect the current evaluation of the method. We may summarise these conclusions as follows:

1. The method is subject to a number of problems, in particular inaccessibility and distortion, but these are not different in kind from those that beset other scientific methods. All can be improved by extending the empirical and theoretical networks, and are ultimately judged by the same criteria of plausibility and consistency.
2. Introspection provides the primary method for data about the experience of thinking, which may be used directly or as a means for making inferences about the process of thinking.
3. The process of thinking may largely be unavailable to consciousness, but introspection can enable a macroscopic description of the structure to be inferred and can suggest hypotheses about the underlying operations.

With regard to the last point, we may bear in mind Nisbett and Wilson's (1977) opinion that verbal reports are based on implicit causal theories about the extent to which a particular stimulus is a plausible cause of a given response. This suggests

that although people may not be able to observe directly their cognitive processes, they will sometimes be able to report accurately about them, when influential stimuli are salient and plausible causes of the response they produce.

Finally, we may refer to some of de Groot's (1966) principles which characterise his introspectively based work. It will be noted that the same principles may be said to apply to the current work. Firstly, the work is directed towards systematic description of cognitive phenomena rather than to strict hypothesis testing; secondly, the experimental settings are more like real life than strictly controlled artificial laboratory conditions; thirdly, protocol coding and interpretation are of crucial importance; fourthly, outcomes will be valuable to the extent that they provide adequate systematic process descriptions. In his 1965 book, de Groot had also specifically addressed the question of reliability. He contends that introspective behaviour is elicited "whenever a person tries to give a truthful answer to the question of what he thinks he is doing, or how, or why he thinks he is doing it"; in addition, the information yielded is generally "quite reliable, in the sense that we actually do rely on it in daily communication and interaction with others". Only if we doubt a person's seriousness or sincerity do we question the reliability of his answers. Bloom (1953) checked the reliability of the method by asking for overt events to be recalled. He found 95 per cent accurate recall of overt, checkable events within two days of the original situation.

Our conclusion must be, then, that, given standardised techniques of implementation and a framework for the subjects' responses, the data elicited from introspective, retrospective techniques (excluding thinking aloud) are reliable and valid. The data, however, may be incomplete to an unknown degree and therefore require inferential responses and conservative interpretation on the part of the experimenter.

9.2.2 The Method of Stimulated Recall

Our particular form of introspection as account gathering by stimulated recall, as already discussed, demonstrates a number of characteristics which will tend to increase reliability and validity of data. These characteristics include standardised implementation, circumscribed content, and external agencies for guidance and stimulating recall (non-directive questioning and videotape playback). In addition, the technique has the great advantage of not interfering in any way with the subject's problem solving thinking process as it occurs.

The first recorded use of a method of stimulated recall in the recent past was by Bloom (1953) who used the technique to uncover students' thoughts during discussions and lectures. He gave the rationale as follows:

"The basic idea underlying the method of stimulated recall is that a subject may be enabled to relive an original situation with vividness and accuracy if he is presented with a large number of the cues or stimuli which occurred in the original situation".

As the stimulus to recall (within 24 hours of the teaching session) Bloom used an audiotape recording. Although the richness of the report protocols indicated great promise for the method, Siegel et al (1963) identify at least three methodological difficulties. Firstly, the period of delay (up to 24 hours) between initial event and playback increases the likelihood of subjects forgetting or suppressing their thoughts during that event. Secondly, the audiotape produces only a portion of the original classroom experience. Cues other than auditory ones could be introduced to stimulate more complete recall. Thirdly, the technique of collecting data from subjects individually is laborious.

While agreeing with Siegel et al's first two points, and having designed the present study to overcome these difficulties by using immediate playback of videotape, as have other workers more recently (Peterson and Clark, 1978), we accept that the

third point is, of course, a problem of time, but not an insurmountable one. Siegel et al (1963) overcome this point in their own research by administering the stimulated recall session to their subjects as a group, stopping the tape at critical points and eliciting written responses. But this technique has two jeopardising factors. Firstly, the identification of critical points should be in terms not only of knowledge conveyed but also of subject activity. Thus true critical points may vary from subject to subject. Secondly, verbal report protocols have the considerable advantage of a high output rate. Having subjects record their accounts in writing is likely to decrease that rate considerably and lose much valuable information. At the same time there is no opportunity to encourage the subject to deepen his account or to prompt responses to special aspects of the stimulus. Stimulated recall sessions with individual subjects are time consuming and laborious, but any diminution of these factors is accompanied by a parallel loss of data.

Apart from the Michigan and McMaster work already described (see section 3.1), another major application of the method of stimulated recall has been in the field of counselling and interpersonal processes. Kagan and Krathwohl (1967) report that it is difficult for a person to introspect and interact with another person in a normal manner simultaneously. They therefore needed "a way of permitting the mind to interact with a situation at one time and to introspect concerning the reaction at another". They concluded that "if we could give a subject enough clues and cues to help him relive the experience, we could explore in depth at a later time various points in the interaction, the thoughts, feelings, changes in thoughts and feelings, and the meaning of various gestures and expressions".

Our current method is closely related to that of Kagan and Krathwohl (1967), who videorecorded the client-counsellor

interview, viewed simultaneously by the experimenter who could then note places of special interest. The client and counsellor were then interrogated separately but simultaneously, watching the same video replay. The interrogations were audiorecorded and transcribed with the original interview. This basic method has been called Interpersonal Process Recall (Kagan, 1977) and it has been shown that "the person is able to recall thoughts and feelings in amazing detail and in depth ... usually a wealth of understanding about some of their underlying motives, thoughts and feelings during the interpersonal transaction could be verbalised by them". It was also found that "the phenomenon could be counted on to work more reliably ... in the presence of a third person who encouraged the viewer to verbalise and elaborate on that which is recalled". It was found that the third person was most effective when actively encouraging the subject by indirect questioning, in the manner of the current study.

Only one methodological problem remains unresolved and unresolvable in any empirical sense. It cannot be established to what degree the subject's retrospective account is distorted by his knowledge of the data subsequently available. If retrospective distortion occurs, to what extent may it be discounted? This problem of retrospective distortion is also addressed by Elstein et al (1978). It is only possible to estimate the true answer to these questions. It is clear that to provide a plausible but inaccurate and distorted account of the thinking processes underlying a complex and lengthy line of medical enquiry which is recorded and available on videotape, would be a complex and skilled operation of rationalisation. The method of stimulated recall allows a check on the plausibility of the subject's account. In addition, the very cognitive complexity and demand on short term and long term memory make a consistent and sustained rationalisation or distortion improbable. Our experience of data collection has shown that subjects have probably not yielded implausible accounts. Neither has there been evidence of attempts to hide errors in thinking or to fail

to differentiate between the thought process as it occurred and a current appraisal of those thoughts in the light of subsequent information. We conclude that distortion of accounts to any jeopardising degree does not occur in the current study. The only identifiable tendency among a small number of subjects is to present their own theoretical interpretation of their thinking processes, in addition to the account itself. This, however, does not jeopardise the validity of data at all and, indeed, may be of positive interest.

Having discussed the theoretical validity and reliability of the method of account gathering by stimulated recall, we may now consider our own implementation of that method, including the quantitative means available for establishing practical reliability and validity.

9.2.3 Selection of Content (Patients)

The current study is based on the use of real patients taken from the hospital wards at the time of each encounter and recall session. Real patients, rather than simulators, were used for a number of reasons. Firstly, our study is primarily a comparison of the performance of three groups of subjects - students, house officers and registrars - rather than a study of the repeated performance of individuals across different content. For this reason, we have relatively large numbers of subjects in each group (22). Secondly, and closely related, our study seeks to identify the generalities of the process of diagnostic problem solving rather than its content specific aspects. A variety of content is appropriate to this aim. That content may vary, within each group of subjects, in terms both of complexity and actual subject matter, although, of course, greater variation between groups of subjects would be considered prejudicial. Thirdly, the use of simulated patients is not yet common in British medical education, although it is certainly known (Maguire et al, 1977). It was considered that the unfamiliarity of a simulation might initially interfere

with the subjects' approaches to the problem. The hospital grapevine would also, almost certainly, transmit news of the simulators' "diagnoses" to prospective subjects. Since each subject saw only one patient, these additive effects might well have been considerable.

Having decided, then, to use real patients, criteria for their selection were determined, remembering that subjects would only be required to take the history. Our criteria were quite simple. The patient should be an in-patient on the given day, on a general medical ward and able to give a comprehensible history (i.e. having no gross communication defect). The patients were not necessarily selected because their disease was especially amenable to diagnosis on the history alone, our study being primarily concerned with the definition of a process and only secondarily with its theoretical, rather than actual endpoint.

Rigorous comparison of cases across groups of subjects is not possible. Appendices 21, 22 and 23 show all diagnoses given in each patient's medical record for cases seen by students, house officers and registrars respectively. Perusal of these shows a wide variety of disorders and systems involved within each group and a spread across groups of cases of carcinoma, endocrinological and haematological disorders, respiratory disease, cardiac problems, gastro-intestinal disorders and so on. Associated tables in Appendix 24 show that the total, range and mean number of diagnoses are not statistically significantly different across groups. This is important since any one patient might have multiple diagnoses. In addition, Table 9.1 has suggested similar patterns across groups in terms of range and mean length of patient interviews which may suggest similar distributions of case difficulty, or complexity across groups.

Finally, one other form of approximate validation of comparability of case content across groups is possible. We may

compare apparent difficulty of cases by comparing subjects' success in reaching the correct diagnosis. If similar success rates are evident, then comparability of difficulty of case content may be inferred. Appendices 21, 22 and 23 show the actual diagnoses given and indicate those which probably cannot be made from the history alone. Appendix 24 (Table 3) shows the numbers of diagnoses which could be made for each group, and the number actually made. Table 4 of that Appendix shows that no statistically significant differences exist between groups on this variable. We may infer valid comparability of case content across groups of subjects.

9.2.4 Other Methodological Aspects: Videorecording and Account Gathering

Possible sources of diminished reliability and validity may be found in the implementation of the research method itself. Firstly, we may identify aspects of the use of video technology, both during the patient interview and the stimulated recall session. On the one hand, the subject is aware of being videotaped, and on the other, he must watch himself on a small screen. Either circumstance may affect the reliability and validity of his responses. In addition, the patient may also be adversely affected by the presence of a camera.

In the event, neither subjects nor patients seemed at all non-plussed by the use of videocamera or playback. Perhaps the domestic television, the cinema, security cameras and photographic equipment have dispelled effects of both fear and novelty. Many patients are quite used to being filmed (or videotaped) and photographed, both in hospital and outside. For whatever reason, most patients displayed no interest whatsoever in the camera or in being 'filmed'. As for the subjects, it is likely that most of them had already had considerable contact with video equipment which is used extensively in both undergraduate and postgraduate medical education, particularly in teaching aspects of interviewing skills (see,

for example, Vaughan and Marks, 1976; Connolly and Bird, 1977; Maguire et al, 1977; Freer, 1978) as well as more clinical skills (Sturgeon, 1979). The experience was therefore unlikely to have been novel for most subjects.

The second methodological aspect concerns the possible invalidating effect of varying treatment of subjects by the experimenter during the account gathering session. This may cause variability in subject responses which, in turn, could invalidate comparison of groups of subjects. Section 9.1 has described the standardised method used for all subjects in all groups. In particular, however, variability of replay periods may contribute to invalidity or unreliability of data. But Tables 9.2, 9.3 and 9.4 suggest that there were no great differences between the three groups of subjects in this aspect of method implementation.

Each subject was asked whether or not the interview with the patient was typical of his normal performance. Out of the 66 subjects only three suggested that it was not typical, in each case due to patient variables. Only four subjects mentioned being aware of the camera, while five said that they normally took less time although the general structure of the interview was typical. Twelve subjects were unable to make a judgment on the matter. These findings are summarised in Table 9.5 and give further support to our evaluation of the data as valid.

Table 9.5 Subjects' Estimates of the Relationship of
Videorecorded Interviews to Usual Performance
(22 subjects per group)

	<u>Typical</u>	<u>Typical but more lengthy</u>	<u>Typical but self- conscious</u>	<u>Don't know</u>	<u>Not typical</u>
Students	15	0	1	4	2
House Officers	14	3	1	4	0
Registrars	13	2	2	4	1

9.2.5 Omission of the Physical Examination

It is well recognised that in most cases history taking from the patient is the most important diagnostic tool (Royal College of General Practitioners, 1972). In an analysis of 80 new patients coming to a medical outpatients clinic, Hampton et al (1975) found that the diagnosis could be made after reading the referral letter and taking the history in 66 cases. The physical examination was useful in only seven patients and the laboratory investigations in a further seven. Rimoldi (1961) using card packs (as described in Chapter Four above) to trace pathways to diagnosis, found that junior students selected more cards (i.e. requested more information) than did senior students who, in turn, selected more cards than did physicians. These differences were due to questions relating to the history and patient interview, not to either the physical examination or laboratory investigations. Rimoldi concludes that "this seems to be in agreement with the suggestion commonly made that, in general, the most significant part of a diagnosis is related to the interview section", and takes this as "evidence that the most striking change in the diagnostic process with increased clinical experience is found in the interview phase". In addition, there is no reason to suppose that data obtained from the physical examination is of a different cognitive order from that elicited as history. Of course, data from the physical examination is of a different quality and nature from that elicited as history, not least because of its direct nature being, at most, mediated by an instrument (stethoscope, ophthalmoscope, sphygmomanometer, etc.) and not via the patient's own cognitive processes. There may be occasions when an item of physical examination is irreplaceable in the diagnostic process, but this may apply to all other types of data also.

It was therefore decided to limit the current study to the central part of the enquiry process and the main diagnostic tool - the history. It is suggested that the findings may be generalised to those situations where the physical examination

also occurs. In terms of the diagnostic thinking process per se, it would seem unlikely that data from different sources should be used in different ways. Likewise, being deprived of the data from a physical examination is unlikely to alter the structure of that basic thinking process, and, in many cases, as the work cited above shows, will not alter its eventual outcome either. Appendices 21, 22 and 23 show that only 21 out of 195 possible diagnoses are identified as being unreasonably difficult to establish on the basis of the history alone. In the event, four of these 21 actually were established by the subject.

9.2.6 Conclusions

Determination of validity and reliability of the method of account gathering by stimulated recall does not present insurmountable problems, although the nature of the proof is largely dependent upon theoretical argument and inference on the basis of theory, unlike the statistical analyses which have been applied to the multiple choice questionnaires in endocrinology and neurology. For such a research instrument, based on untrained introspection, the manner of implementation of the method and a standardised approach towards subjects becomes crucial. Such aspects of the design of the instrument as are amenable to empirical study have been discussed, but, again, the order of analysis is less exacting than that applied to our complementary research instrument. It is, nonetheless, appropriate and reasonably sufficient.

The validity and reliability of data have been discussed and determined. It is incumbent upon the experimenter not to push interpretation of that data beyond its point of reasonable parsimony and caution.

We may now proceed to discuss the methods of data analysis applied to the transcribed accounts.

9.3 Data Analysis

Transcripts were initially analysed by a process of content

analysis. The results of this process are frequencies of defined categories of response. These were subjected to various non-parametric statistical analyses. The interpretation of protocols, such as the accounts of the present study, involves certain considerations. These are discussed in Appendix 25. The method of content analysis itself and the derivation of the categories and indicators for this study are described in Appendix 26. The categories and indicators are described later in section 10.1. The categories and raters using them must be subjected to tests to establish their reliability and validity. Appendix 27 describes the methods applied in this study and presents results which show that reliability and validity of categories and raters are established.

The types of statistical analysis which are necessary and possible from hypothesis to hypothesis are very similar, by virtue of the similar nature and structure of the data across all hypotheses. Therefore, instead of describing the precise statistical analysis for each hypothesis separately, we shall discuss the two statistical tests used and describe the purpose of that use. Section 6.5 above presents a detailed discussion of the criteria for selection between parametric and non-parametric tests. Although the data from content analysis meet many of the necessary criteria for justified application of parametric statistical tests, there is one characteristic in particular of those data which has determined the use of non-parametric statistical tests. This characteristic is the often very small numbers which must be analysed. On occasions (where, for example, $N < 3$) these figures have become even too small for non-parametric statistics and data is assessed by inspection and tabulation only. Hypothesis 10 is the only exception, providing data amenable to parametric analysis. However, two non-parametric tests are used; these are the χ^2 (chi square) test for K independent samples, and the Kruskal-Wallis one-way analysis of variance. Table 9.6 shows

which hypotheses and categories are subjected to which of these tests. We may consider in turn the rationale for application of each of these tests as applied to the present data.

Table 9.6 Hypotheses and Categories Subjected to the χ^2 Test for k Independent Samples and the Kruskal-Wallis One-Way Analysis of Variance, Respectively

<u>Hypothesis</u>	<u>Categories</u>	<u>χ^2</u>	<u>Kruskal-Wallis</u>
10	a, b	Parametric two-way ANOVA	
11*	a, b, c	✓	
12	a, b, c	✓	
13	a, b		✓
14	a, b, c		✓
	1,2 or 3 of a,b,c	✓	
15	a, b, c		✓
16	a, b		✓

* Parametric one-way analysis of variance also performed on data of length of first replay periods.

9.3.1 The χ^2 Test for Independent Samples

When data are ranged in the form of contingency tables, χ^2 provides an appropriate test of independence of the two variables (Ferguson, 1966). When categories per hypothesis are taken as levels of one variable and groups of subjects as levels of the other so that data may be seen as paired observations, the independence of these may be determined. The χ^2 test may be used satisfactorily when frequencies constitute the data of research (Seigel, 1956). .

χ^2 may only be used where fewer than 20 per cent of cells have expected frequencies of less than 5 and no cell has an expected frequency of zero. χ^2 applied to data from Hypothesis 16

would have violated the former of these conditions. χ^2 was calculated according to the formula:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

9.3.2 The Kruskal-Wallis One-Way Analysis of Variance by Ranks

This test is useful for deciding whether k independent samples come from different populations. It tests the null hypothesis that the samples come from the same population or from identical populations with respect to averages. The test may be applied either for small ($k = 3, n_j < 5$) or large ($k > 5, n_j > 5$) samples. The value of H (the test statistic) is defined by the formula:

$$H = \frac{12}{N(N+1)} \sum \frac{R_j^2}{n_j} - 3(N+1)$$

where R_j is the sum of ranks in the j th sample. The statistic has $k-1$ degrees of freedom and is based on the conversion of scores to ranks (see Seigel, 1956 for a detailed account).

In the present analysis the Kruskal-Wallis H statistic is used to test the hypothesis of no difference between groups, or no difference between categories after having shown independence of groups and categories by means of the χ^2 statistic.

No other statistical test was necessary.

9.4 Summary

The implementation of the account gathering research method in this study is described and discussed. The validity of the method is established by means of theoretical analysis of the rationale of introspectionist methods and the method of stimulated recall as a special case of introspection. The selection of content (patients) for the present study is described and discussed and the validity of that content for the three groups of subjects established. The methodological validity of videorecording and this special instance of account gathering is established and the rationale for omission of the

physical examination described and discussed. The two stages of data analysis are discussed. Firstly, the method and rationale of content analysis is described including establishment of methodological and content validity and reliability. Statistical analysis of the results of the content analysis is described.

CHAPTER TEN

The Account Gathering Study 2: Results

This chapter presents the results of analyses applied to the transcripts derived from the videotape stimulated recall sessions with students, house officers and registrars. These will be presented in relation to the process of content analysis and the research hypotheses, described in Chapter Nine. Full interpretation and discussion of the results in relation to the research hypotheses stated in section 5.1 above is reserved for Chapter Eleven.

10.1 Content Analysis: Categories and Indicators

Appendix 26 describes the method of deriving categories and indicators for the content analysis of transcripts. The figure in Appendix 26 represents the circular, repeating process which begins with the research hypotheses and ends any number of cycles later when the system of categories eventually proves itself suitable to hypotheses and data alike. Two of the categories (pre-diagnostic and diagnostic interpretations) have already been mentioned because of the particular form of reliability study which was applied to them. We may now define these and all other categories, with their indicators, as they emerged from the process of content analysis. The categories and indicators define the research hypotheses from which they are ultimately derived; the relevant hypothesis is therefore identified by number for each set of categories. It will be remembered that only Hypotheses 10 to 16 were treated in this manner. Hypotheses 17 to 20 do not have results which can be presented separately for discussion. These hypotheses, therefore, will not be considered again until Chapter Eleven. Categories and indicators for each of Hypotheses 10 to 16 emerged and were defined as follows.

10.1.1 Hypothesis 10 Categories and Indicators

Category (a) Pre-diagnostic interpretation of clinical information

Definition Any term, phrase or statement which indicates that

S has made some active interpretation of the clinical information available to him WHERE the result of this activity is not sufficiently specific to constitute a possible diagnosis.

Indicators (i) The resultant interpretation may be at a relatively undefined level, where no clearly specified pathological process is mentioned in terms of descriptive, clinical, metabolic, morbid anatomical, infective, inflammatory, biochemical, genetic or psychiatric pathophysiological processes:

e.g. "Something wrong with".

"Myocardial problem".

"Psychiatric disturbance".

(ii) The resultant interpretation may be at a relatively defined level where a pathophysiological process is indicated in the above terms, but the interpretation would require further specificity before constituting an acceptable final diagnosis:

e.g. "A metabolic abnormality".

"Anaemia caused by blood loss".

"Acute hepatitis".

Category (b) Diagnostic interpretation of clinical information

Definition Any term, phrase or statement which indicates that S has made some active interpretation of the information so far available AND where a pathophysiological process is indicated with a greater degree of specificity than for Category (a) sufficient for a diagnosis.

Indicators This need not be the greatest degree of specificity possible but should seem to be sufficient as a diagnosis in a normal clinical context:

e.g. "Carcinoma of the pancreas".

"Acromegaly".

"Steroid induced Cushing's".

10.1.2 Hypothesis 11 Categories and Indicators

N.B. All categories are with reference to S's interpretative or evaluative response to information elicited from the patient during the first replay period only (which response may be reported either during the first stop or more retrospectively during a later stop).

Category (a) as Category (a) for Hypothesis 10

Category (b) as Category (b) for Hypothesis 10

Category (c) Judgment of need for further general or clarifying enquiry, not stemming from either pre-diagnostic or diagnostic interpretations

Definition Where S enquires further about the patient's symptoms, signs, etc. for clarification OR where S seeks to clarify the patient's statement.

N.B. Not, where S is seeking a particular piece of information based on his own expectations.

Indicators "I was asking how the pain affected him".

"I asked her how she didn't feel well".

"... find out what the patient means ..."

"... trying to get him to tell me what was the matter with his ankle ..."

10.1.3 Hypothesis 12 Categories and Indicators

Category (a) as Category (a) for Hypothesis 10

Category (b) as Category (b) for Hypothesis 10

Category (c) Expecting, searching for or planning to search for specific features (symptoms, signs, tests, etc.) of disease or treatment of disease

Definition Where S shows expectation of certain clinical information or considers certain features of disease likely or possible in the patient, given the

information already elicited.

N.B. Exclude specific questions which form part of the routine, systematic review.

Indicators "If we investigated the patient, I'd imagine we'd find X".
"I'm asking these questions because the patient may have X as part of her syndrome".
"I was thinking it might be diagnosis Y, so I went through other things typical of that".
"I wanted to know whether the pain was due to X or Y, so I'm asking those questions".
"I asked questions about other possible diagnoses and conditions".

10.1.4 Hypothesis 13 Categories and Indicators

Category (a) Reinterpretation of clinical information, when no new information has been added

Definition Where an array of clinical information which has already been interpreted in some way becomes amenable to a new (altered or additional) interpretation because of a change in S's own thinking and not because new information has been added to the array. The new interpretation may or may not be related to the old one(s), i.e. it may be a progressive refinement or specification of the old interpretation(s) or it may be a completely new interpretation.

Indicators "It was creeping into my mind/struck me/flashed through my mind that he may have diagnosis X" (when no new information has prompted this).
"I suddenly saw that symptoms X and Y were related/separate".
"I had finished, but I just asked that question because I suddenly thought she might have diagnosis X".

"I started thinking again why it had all started and I thought of diagnosis X".

Category (b) Reinterpretation of clinical information arising from the addition of new information

Definition Where an array of clinical information which has already been interpreted in some way becomes amenable to a new (altered or additional) interpretation because of the addition of new information to the array. The new interpretation may or may not be related to the old one(s), i.e. it may be a progressive refinement or specification of the old interpretation(s) or it may be a completely new interpretation.

Indicators "Symptom X now suggests that it may be diagnosis Y".
"Then the patient told me he had symptom X, which made me think it might be something completely different from what I'd been thinking of".
"I'd thought of diagnosis X, but when I asked further questions, I realised that diagnosis Y was the case".

10.1.5 Hypothesis 14 Categories and Indicators

Category (a) Active confirmation of an interpretation

Definition Where S feels that the selected interpretation is confirmed as an actual diagnosis.

Indicators "My conclusion is that she's suffering from diagnosis X".
"I'm convinced by now that it's diagnosis X".
"I know one diagnosis - she probably had X".
"He's describing syndrome X, which is A + B + C due to E".

Category (b) Active elimination of an interpretation *

Definition Where S eliminates an identified interpretation

because of contrary evidence or positive lack of necessary evidence.

- * The complementary category "passive elimination" where interpretations decay, are forgotten or eliminated by implication cannot, by definition, be determined from the data with any reliability. It is therefore omitted.

Indicators "I'd pretty well dismissed diagnosis X because he hadn't got symptom Y".
"I asked about X because it can cause diagnosis Y, but she hasn't got X".
"I realised it wasn't diagnosis X".
"If she'd had symptom X it would have been diagnosis X, but I think it was diagnosis Y".

Category (c) Postponement of either confirmation or elimination of a possible interpretation with or without stated differential likelihoods

Definition Where an identified possible interpretation is neither confirmed nor eliminated by S, but is left under postponed judgment with or without S giving an estimate of its likelihood.

N.B. This category can be designated only after having read the entire transcript since interim postponements of judgments are not indicative of this category.

Indicators Where the following are S's final opinion:
"It was disease type X, with diagnosis Y high on the list".
"She's the right category for disease X, but I don't know any questions that could diagnose that".
"I'd put my money on diagnosis X if I had to, but I couldn't really say with a great deal of certainty that that's what happened to her".

Note: The above categories cannot be seen as absolute.

Categories (a) and (b) may be subject to reversal given new information. These categories merely describe cognitive operations which may be present during the whole thinking process. It is psychologically untenable to suggest that a possible interpretation, once made, can be totally expunged or confirmed without chance of later doubt or is not capable, when eliminated, of being later reidentified. Our interest is in the presence or not of the cognitive operation.

10.1.6 Hypothesis 15 Categories and Indicators

Category (a) Patient - determined interview structure

Definition Where the course of the interview as directed by S is determined by, or follows on from, the flow of information as presented by the patient.

Indicators

- (i) NOT where S merely allows the patient to talk without interruption, which is stylistic rather than structural.
- (ii) NOT where S merely seeks clarification or elaboration of information given to the patient.
- (iii) "If a system came up, I dealt with it there instead of waiting for the systematic enquiry".
"I decided to do the CVS there because it was relevant to what she just mentioned".
"I went on to the gut because that's what he seemed interested in".

Category (b) Subject - determined interview structure

Definition Where the course of the interview is determined by the subject's requirement actively to test his interpretations of the clinical information.

Indicators "I was thinking in terms of diagnosis X, so I asked about symptom X".

"I was looking for symptoms X, Y and Z".

"I'm still plugging along the thyroid side".

"I'm excluding some of the pointers".

Category (c) Logically - determined interview structure

Definition Where the subject conducts, or attempts to conduct, the interview according to a routine format as defined by the standard (taught) clinical history or any of its component parts. Exclude the presenting complaint since all subjects begin with this area. Define the standard (taught) clinical history as follows:

1. Presenting complaint.
2. History of the present complaint.
3. Symptomatic survey (by systems):
 - a) Cardiovascular
 - b) Respiratory
 - c) Locomotor
 - d) Genito-urinary
 - e) Gastro-intestinal
 - f) Neurological
 - g) General.
4. Past medical history.
5. Family history.
6. Social history.
7. Drug survey.

Indicators "This is just routine/general enquiry".

"He (the patient) was getting my systems mucked up, so that had to wait".

"I suspected X right at the beginning, but I stored it away because it had to be asked later in the systematic enquiry".

"I knew all about the history of the present complaint, so I started on the old list".

Note: In essence, the course of the interview is always subject determined, the above categories identify the factors which influence S's decisions about the course of the interview at various points during its progress.

10.1.7 Hypothesis 16 Categories and Indicators

Category (a) Failure to make specific enquiry

Definition Where S identifies, in retrospect, his own failure to make relevant, specific enquiry concerning the patient's problem, symptoms, signs, etc.

Indicators "I should have gone into that symptom in more depth, but I forgot".
"Symptoms kept cropping up all over the place that I hadn't thought to ask about".
"I should have asked X (with regard to the patient's complaint)".

Category (b) Failure to make general enquiry

Definition Where S identifies, in retrospect, his own failure to make sufficient routine, general or screening enquiry.

Indicators "This is just a superficial social history, it might be worth pursuing it a bit more".
"I could have asked a lot more questions about this system, but I tend to forget them unless they seem necessary".
"That's a terrible GI history. I hadn't got my thoughts marshalled sufficiently to rap out all the direct questions".
"(Under neurological system) I didn't ask about headache".

Note: These categories indicate S's failure at the time, not a failure of knowledge generally, but a failure to "think in action".

10.2 Statistical Analyses

Section 9.3 above discusses the two major statistical tests applied to the results of content analysis. These results are frequencies representing the number of subjects per group who displayed evidence of a given category, with the exception of categories 10(a) and 10(b) which concern the total number of responses of each category made by each subject. The two major statistical tests applied are the χ^2 test for k independent samples and the Kruskal-Wallis one-way analysis of variance by ranks. It will be noted from Table 9.6 that Hypothesis 11 is also subject to a parametric one-way analysis of variance. Since results are presented hypothesis by hypothesis, these matters will be discussed further below. Because of the essential interrogative nature of Hypotheses 17 to 20, these are not subject to any form of statistical analysis and will not be considered again until Chapter Eleven.

All χ^2 results will be presented as contingency tables, with observed and expected frequencies, marginal totals and the derived value of χ^2 .

Kruskal-Wallis' H will be presented with values of all n's, k and N, degrees of freedom and correcting factors where necessary. Because of the limited nature of this information, Appendix 29 presents the raw data of content analysis in terms of frequencies of incidence per group per category.

10.2.1 Hypothesis 10 Parametric Two-Way Analysis of Variance

Appendix 30 presents the raw data of content analysis using categories 10(a) (pre-diagnostic interpretations) and 10(b) (diagnostic interpretations). It may readily be seen that these data are amenable to parametric statistical analysis not contravening any of the criteria for such analyses discussed in section 6.5 above. Accordingly, a two-way analysis of variance was applied. Table 10.1 presents the summary of that calculation. It may be observed that the interaction term is not statistically significant, suggesting independence of the

Table 10.1 Two-Way Analysis of Variance Summary Table
Hypothesis 10 Categories (a) and (b)

<u>Source of variation</u>	<u>Sums of squares</u>	<u>Df</u>	<u>Variance estimate</u>
Rows (groups)	13.6	2	6.8
Columns (categories)	251.3	1	251.3
Interaction	3	2	1.5
Within	1583.9	126	12.6
Total	1851.8	131	
$F_i = 0.1190$	$F_r = 0.5397$	$F_c = 19.9444$	
(df. 2,126)	(df. 2,126)	(df. 1,126)	
NS	NS	$p < .01$	

groups and categories dimensions. In other words, relative use of categories does not alter with status of the subject. The main effect of rows (groups) does not reach a statistically significant value either, indicating, as we may expect on the basis of the non-significant interaction term, that groups of subjects do not differ in their relative use of the categories. However, the main effect for columns (categories) does achieve statistical significance, from which we infer a difference in relative use of category (a) and category (b) responses. Table 10.2 summarises the raw scores on each of these for each group. It is clear that category (a) responses (pre-diagnostic inter-

Table 10.2 Summary of Raw Results for Categories 10(a)
and 10(b) (n = 22 for each group)

Group	10(a)			10(b)			Difference in totals 10(a)-10(b)
	Total	Mean	S D	Total	Mean	S D	
Students	154	7	4.56	91	4.1	2.05	+63
House Officers	164	7.4	3.87	96	4.4	3.05	+68
Registrars	165	7.5	4.57	114	5.2	2.28	+51

pretations of information) predominate over category (b) responses (diagnostic interpretations of information) for all groups.

10.2.2 Hypothesis 11 Parametric One-Way Analysis of Variance and χ^2

This analysis concerns subjects' responses to information elicited during the first replay period only. It is therefore necessary to establish the length of that period for each group and the lack of statistically significant differences between groups on this variable before we may interpret the results of content analysis on the basis of the three associated categories. Parametric one-way analysis of variance is quite appropriate for this purpose since the data do not violate any of the criteria set out in 6.5 above.

Table 10.3 summarises the length (in seconds) of the first replay period for each group. No indication of whether or not the main complaint was elicited during this time is given because of the difficulty of defining exactly what constitutes the main complaint in any one case, and the illogicality of disregarding information to which the subject must respond. Should any symptom relevant to one of the major diagnoses be accepted as a main complaint? Should any symptom or sign or piece of clinical information relevant to any of the patient's diagnoses be accepted? What reason is there to discount any information which the patient gives, regardless of its status with regard to the patient's diagnoses, since the subject must evaluate all information in order to decide what to disregard and what to retain? Barrows et al's (1978) formulation is therefore eschewed as inappropriate. It is hypothesised here that all information must receive response from the subject, regardless of its relationship to the actual diagnosis, for it is that relationship which the subject must determine. We therefore provide only information concerning the length of the first replay period and not its content which may or may not have been relevant to the actual diagnosis. It can be seen from Table 10.3 that the first replay period varied in

Table 10.3 Length (in seconds) of the First Replay
Period (n = 22 for each group)

<u>Group</u>	<u>Mean</u>	<u>Range</u>	<u>S D</u>
Students	49.05	15 - 73	16.82
House Officers	41.12	13 - 82	22.02
Registrars	48.89	24 - 75	14.39

mean length from 41 to 49 seconds and actually ranged from 13 to 82 seconds.

Table 10.4 shows a non-significant value of F, indicating no statistically significant difference between length of first replay period for students, house officers and registrars. Having established these features, we may go on to consider the groups' responses in relation to categories (a), (b) and (c).

Table 10.4 One-Way Analysis of Variance Summary Table
for Length of First Replay Period

<u>Source of variation</u>	<u>Sums of squares</u>	<u>Df</u>	<u>Variance estimate</u>
Between	731	2	365.5
Within	19542	63	310.19
Total	20273	65	F = 1.1783 (df. 2,63) NS

Table 10.5 summarises the observed frequency of response in each category. Inspection of this table reveals that some responses were made in each category. However, from this

Table 10.5 Observed Frequency of Response in Categories
11(a), (b) and (c) for Each Group (n = 22 per group)

Group	Category			Total
	11(a)	11(b)	11(c)	
Students	10	7	9	26
House Officers	17	8	4	29
Registrars	15	7	6	28
Total	42	22	19	83

table alone, we cannot tell whether all subjects made a response in at least one of the categories, which is our concern. Table 10.6 gives this information, showing how many members of each group made no such responses, a response in one of the categories only, or in two or in all three. In other words, responses of more than one type may be made to the same information.' It can be seen from Table 10.6 that only two students and one house officer failed to make a response in any of the three categories, while most subjects made a response of one type only (Table 10.5 would suggest that category (a) responses were more common than category (b) ones). However, almost one third (19) of the subjects made responses of two types, while only one subject actually made responses in all three categories. It will be noted that the values of the frequencies in each cell preclude any useful statistical analysis.

The precise nature of the single and dual responses is not of great value in illuminating the subjects' cognitive processes, since their responses are more likely to be dependent upon the information elicited rather than their own tendencies to think in certain ways. What is of generalisable importance is that a subject can and does respond to clinical information

Table 10.6 Frequency of Single and Multiple Responses to Information Elicited During the First Replay Period (n = 22 per group), Categories 11(a), (b) and (c)

<u>Group</u>	<u>Number of Categories of Response</u>			
	0	1	2	3
Students	2	15	4	1
House Officers	1	13	8	0
Registrars	0	15	7	0
Total	3	43	19	1

on one or many levels as reflected by the categories. However, remembering the possibly partially information-determined nature of responses, it may be interesting to consider what these actually were. Table 10.7 gives this information. It will be seen for all groups that category (a) alone is the most common single response, while category (b) responses are the least common. Of the combined categories, (a) plus (b) form the only pair used to any appreciable extent. It must be emphasised that no strong conclusions can be drawn from Table 10.7, but it may be useful in supporting inferences made on more solid grounds.

Table 10.7 Frequencies of Subjects' Responses in Single and Combined Categories 11(a), (b) and (c) (n = 22 per group)

<u>Groups</u>		<u>Categories</u>						
		None	(a)	(b)	(c)	(a)+ (b)	(a)+ (c)	(b)+ (c)
Students		2	6	3	6	2	1	1
House Officers		1	9	2	2	7	1	0
Registrars		0	8	2	5	5	2	0
Total		3	23	7	13	14	4	1

Inspection of Table 10.7 seems to suggest that the pattern of responses is similar for all groups, but the data are not amenable to statistical analysis. However, a χ^2 test may be applied to determine whether or not the variables (group membership and response category) are independent. Table 10.8 demonstrates that this is the case. We may conclude that category of response to initial information is not determined by status of the subject responding.

<u>Table 10.8 Summary of χ^2 Test to Determine the Independence of the Variables: Status and Response Category for 11(a), (b) and (c)</u>				
Group	Category			
	11(a)	11(b)	11(c)	Total
Students	10 (13.1)	7 (6.9)	9 (5.9)	26
House Officers	17 (14.7)	8 (7.7)	4 (6.6)	29
Registrars	15 (14.2)	7 (7.4)	6 (6.4)	28
Total	42	22	19	83
$\chi^2 = 4.1349$ (df.4) NS				

10.2.3 Hypothesis 12 χ^2

Although a χ^2 test is applied to Hypothesis 12 data, this is really a matter of formality since, as Table 10.9 indicates, frequency of response in each category is 100 per cent in seven of the nine cells and only one count is missing in each of the remaining two cells. It is not surprising, therefore, to find that the variables are independent.

10.2.4 Hypothesis 13 Kruskal-Wallis

Table 10.10 shows the frequency with which different types of reinterpretation of clinical information are used by each group of subjects. It will be noted that the figures are not

Table 10.9 Summary of χ^2 Test to Determine the Independence of the Variables: Status and Frequency of Response per Category for 12(a), (b) and (c)

Group	Category			
	12(a)	12(b)	12(c)	Total
Students	22 (21.5)	21 (21.5)	22 (21.9)	65
House Officers	22 (21.9)	22 (21.9)	22 (22.2)	66
Registrars	21 (21.5)	22 (21.5)	22 (21.9)	65
Total	65	65	66	196
$\chi^2 = 0.0666$ (df.4) NS				

Table 10.10 Frequency of Responses per Group in Categories 13(a) and 13(b) Separately and Conjointly and Number of Subjects Making Responses in Neither Category
(n = 22 per group)

Group	Category				Total
	Neither	13(a)	13(b)	Both	
Students (per cent)	10 (45)	0 (0)	9 (41)	3 (14)	22 (100)
House Officers (per cent)	6 (27)	2 (10)	10 (45)	4 (18)	22 (100)
Registrars (per cent)	10 (45)	0 (0)	8 (37)	4 (18)	22 (100)
Total (per cent)	26 (39)	2 (3)	27 (41)	11 (17)	

amenable to statistical analysis. Inspection of the data reveals that the most common response is represented by category 13(b) which indicates reinterpretation of clinical information on the basis of new information being added to the array. It is much less common in the sample for reinterpretation to occur in the absence of new information (13 out of a possible 66 occasions which is approximately 20 per cent of occasions). It is important to note, however, that 26 (39 per cent) of subjects make no reinterpretations of the clinical information at all. This allows us to make the important inference that the first interpretation of any piece or pieces of clinical information which the subject makes is not necessarily subject to alteration or modification. This primary interpretation may or may not be correct, but in 39 per cent of the samples it persisted throughout the diagnostic thinking process. Given this finding, it is worthwhile determining whether or not the response pattern is similar across all subjects. We may, therefore, apply a Kruskal-Wallis one-way analysis of variance by ranks to the raw data of Table 10.10. The results of this test are unequivocal. Where $k = 3$, $n_1 = 4$ and $n_2 = 4$ and $n_3 = 4$ and $N = 12$, $H = 0$ with 2 df. This, clearly, is not a statistically significant value. We may conclude that the three groups of subjects do not differ in their response patterns as measured by categories 13(a) and (b).

10.2.5 Hypothesis 14 χ^2 and Kruskal-Wallis

Hypothesis 14 suggests that no difference will be found between students, house officers and registrars in frequency of response for each of the three categories which reflect strategies for selection among competing interpretations of clinical information. We may therefore commence by testing this hypothesis directly, by the Kruskal-Wallis one-way analysis of variance by ranks comparing groups. Where $k = 3$, $n_1 = 3$, $n_2 = 3$ and $n_3 = 3$, and $N = 9$, $H = 0.7923$, corrected for ties, this value becomes 0.8126. This value is not statistically significant. We may therefore conclude that groups of subjects do not differ in frequency of use of each of the categories of Hypothesis 14 (see Appendix 29 for raw data).

However, although frequency of use of categories does not differ from group to group, it may be, as with Hypotheses 11 and 13, that groups differ in their use of combinations of categories. It would be a reasonable contention that in the present case such differences or similarities are less dependent upon the information elicited and more dependent upon the subjects' own cognitive styles ("self-consistent and enduring individual differences in cognitive organisation and functioning, ... in general principles of cognitive organisation ... and ... various self consistent idiosyncratic tendencies ...". Ausubel et al, 1978). This is unlike the case of Hypothesis 11 categories, where style of use is presumed to be primarily dependent upon the clinical information presented. To test this aspect of Hypothesis 14, we may count the frequency of use of one, two or three of the categories in subjects of each group and determine whether or not groups and strategies of category use are independent variables by means of a χ^2 test. Table 10.11 presents results of this test and demonstrates that, indeed, these variables are independent ($\chi^2 = 1.6674$, df.4, NS).

Table 10.11 Summary of χ^2 test to Determine the Independence of the Variables: Status and Strategies of Use of Categories 14(a), (b) and (c) (n = 22 per group)

<u>Group</u>	<u>Number of Categories Used</u>			<u>Total</u>
	<u>1</u>	<u>2</u>	<u>3</u>	
Students	5 (6.3)	12 (12.3)	5 (3.3)	22
House Officers	7 (6.3)	13 (12.3)	2 (3.3)	22
Registrars	7 (6.3)	12 (12.3)	3 (3.3)	22
	19	37	10	66

$\chi^2 = 1.6674$ (df.4) NS

Such independence of variables suggests that all groups adopt the same strategy, however, it does not tell us whether or not the possible strategies are differently used. We may, therefore, apply a Kruskal-Wallis test to determine whether or not such differences are apparent. Where $k = 3$, $n_1 = 3$, $n_2 = 3$ and $n_3 = 3$, and $N = 9$, $H = 6.8$, when corrected for ties, $H = 6.9772$. For this value $p < .01$. We may therefore conclude that there is a statistically significant difference between subjects' use of either one, two or three categories. Inspection of Table 10.11 shows that the use of two categories is the most common strategy. Table 10.12 may assist discussion of this finding by presenting the actual observed frequencies of subjects' responses in all single and combined categories. However, inspection of this Table does not yield any immediately interpretable finding.

Table 10.12 Frequency of Subjects' Responses in Single and Combined Categories 14(a), (b) and (c)
(n = 22 per group)

<u>Groups</u>	<u>Categories</u>						
	(a)	(b)	(c)	(a) + (b)	(a) + (c)	(b) + (c)	(a) + (b) + (c)
Students	3	1	1	7	3	2	5
House Officers	2	1	4	5	2	6	2
Registrars	2	0	5	5	2	5	3
Total	7	2	10	17	7	13	10

It is untenable, given the non-significant value of χ^2 for comparing groups, to suggest that strategy of category use is dependent upon the subject's knowledge base. It may, however,

be inadmissible to suggest that logical considerations are at work, since although the use of more than one category by any one subject does not imply that the subject has a dual or triple evaluation of the same interpretation of the clinical data, he may be dealing with two or three interpretations of the same data and evaluate each differently, but his interpretation of one may have logical consequences or be dependent upon his interpretation of others. Results relevant to Hypothesis 14 thus demonstrate flexibility of thought in all groups of subjects.

10.2.6 Hypothesis 15: Kruskal-Wallis

This hypothesis suggests that no differences will be observed between groups in their use of categories 15(a), (b) and (c). Kruskal-Wallis one-way analysis of variance by ranks indicates the validity or invalidity of this hypothesis. Where $k = 3$, $n_1 = 3$, $n_2 = 3$ and $n_3 = 3$, and $N = 9$, $H = 0.6057$, when corrected for ties, $H = 0.5501$. This value is not statistically significant. We may therefore conclude that Hypothesis 15 is substantiated. However, although groups do not differ in their use of categories (a), (b) and (c), the relative frequency of use of each category may differ across all groups. A Kruskal-Wallis analysis can also illuminate this possibility. Where k , n 's and N are as above, $H = 31.41$, when corrected for ties $H = 34.5811$. This value is statistically significant ($p < .01$). Inspection of the observed frequencies per category (see Table 10.13) shows that this significant value derives from the infrequent use of category (a) and very frequent use of categories (b) and (c). This significant value of H , given its derivation and the non-significant value of H for comparison of groups, indicates that all subjects conduct the clinical interview according to the demands of either their own interpretations of the information or the requirements of some standard form of interview structure and content or, as is usually the case, according to some combination of both criteria. The value of 6 for students in category (a) probably only reflects either a lack of current expertise in the

Table 10.13 Observed Frequency of Use of
Categories 15(a), (b) and (c) (n = 22 per group)

<u>Groups</u>	<u>Categories</u>			
	(a)	(b)	(c)	Total (max. 66)
Students	6	21	21	48
House Officers	1	20	21	42
Registrars	1	21	18	40
Total	8	62	60	130

interpersonal aspects of interviewing or a lack of sufficient command of knowledge to maintain the pace and direction of the interview throughout. However, the results indicate quite clearly the subject's need to control the flow of information according to his own designs. The Baconian ideal of thorough data collection followed by interpretation would appear not to be followed by these subjects and to be questionable in the light of these results.

10.2.7 Hypothesis 16: Kruskal-Wallis

Table 10.14 shows the frequency of observed response in categories (a) and (b). It can be seen that the incidence of these responses is not very great, ranging from four per cent to 54 per cent across all cells. Application of Kruskal-Wallis one-way analysis of variance by ranks where $k = 3$, $n_1 = 2$, $n_2 = 2$, $n_3 = 2$ and $N = 6$, $H = 0.2857$, when corrected for ties $H = 0.2941$. This value is not statistically significant. We may therefore conclude that the three groups do not differ in their use of the two categories. Although no statistical test is available, it is clear that category (a) responses are more common than category (b). In other words, all groups retrospectively identify failure to make a specific enquiry more often than they

Table 10.14 Frequency of Observed Responses in
Categories 16(a) and (b) for all Groups (n = 22 per group)

<u>Groups</u>	<u>Categories</u>	
	(a)	(b)
Students (per cent)	12 (54)	3 (14)
House Officers (per cent)	8 (36)	1 (4)
Registrars (per cent)	4 (18)	4 (18)
Total (per cent)	24 (36)	8 (12)

identify failure to make a general or routine enquiry. However, the values of the observed frequencies in each cell are relatively low. Even when compared with an arbitrary theoretical frequency of 14 (approximately 64 per cent frequency) the values are statistically significantly lower than this (see Table 10.15).

Table 10.15 Comparison of Observed Frequencies in
Categories 16(a) and (b) with Arbitrary Theoretical
Frequency of 14 (c. 64 per cent) (n = 22 per group)

<u>Category</u>	<u>O</u>	<u>E</u>	<u>(O-E)²</u>	<u>$\frac{(O-E)^2}{E}$</u>	
(a)	12	14	4	0.2857	$\chi^2 = 9.9999$ df.2 p < .01
	8	14	36	2.5714	
	4	14	100	7.1428	
(b)	3	14	121	8.6428	$\chi^2 = 27.857$ df.2 p < .01
	1	14	169	12.0714	
	4	14	100	7.1428	

Our conclusion must be, then, that while some subjects fail to "think on their feet", the total number of these is not statistically significant even when compared with a relatively low arbitrary figure. The largest frequency occurs for students, in category (a), reflecting, possibly, that an easily manipulable or retrievable knowledge structure has, as yet, to be developed.

No other statistical tests or tables are considered necessary for Hypotheses 10 to 16.

10.3 Summary

Categories and indicators are described which emerged from the process of content analysis in relation to Hypotheses 10 to 16 which they define and from which they are ultimately derived. Statistical analyses of the results of content analysis are described for each hypothesis and results of those analyses presented. The major tests applied are the χ^2 and Kruskal-Wallis one-way analysis of variance by ranks, both of which are non-parametric. The essential interrogative nature of Hypotheses 17 to 20 renders statistical analysis inappropriate. Discussion of results for Hypotheses 10 to 16 and of findings relevant to Hypotheses 17 to 20 follows in Chapter Eleven.

CHAPTER ELEVEN

The Account Gathering Study 3: Discussion of Results

In Chapter Ten were presented the results of content analysis and subsequent statistical analysis pertaining to Hypotheses 10 to 16. A discussion is now presented of those results and of findings pertaining to Hypotheses 17 to 20 also. Where hypotheses are substantially related, they are discussed in conjunction with one another. The discussion in this chapter is limited to the results of the account gathering study alone. In Chapter Twelve this and the discussion of results of the questionnaire study are combined to form a unified view of the diagnostic thinking process, drawing upon the psychological literature for its explanatory power. Meanwhile, the discussion of the present chapter is limited to Hypotheses 10 to 20, and considers each in turn in the light of the Michigan and McMaster studies as appropriate (see 3.1.1 and 3.1.2), and culminates in a discussion of the interpretative value of the present study for such current descriptions of the diagnostic thinking process.

11.1 Overview of the Issues for Discussion

The discussion of results of the account gathering study is intended to define more closely both previously identified and new variables within the diagnostic problem and thinking process. In addition, a more useful and complete analytical framework than hypothesis generation and testing is sought to explain in psychological terms the cognitive processes of the clinical problem solver.

11.2 Hypothesis 10

- (a) "Students, house officers and registrars will make multiple pre-diagnostic and diagnostic interpretations of clinical information as it is progressively elicited during the clinical interview.
- (b) No differences will be observed between students, house officers and registrars in the relative use of pre-diagnostic and diagnostic interpretations of clinical information during the entire diagnostic thinking process."

Barrows et al (1978), Barrows (1976) and Neufeld et al (1976) refer to 'broad' or 'general' and 'specific' hypotheses, although despite giving examples no definition of these terms seems to be available. Barrows and Bennett (1972) suggest that students and house officers generate precise and specific hypotheses, whereas good clinicians tend to keep theirs broad and vague, allowing them to be shaped by the data. However, Barrows et al (1978) report that physicians' early hypotheses may be either specific or general. Overall, the current McMaster view is somewhat unclear (see 3.1.2 above). Elstein et al (1978) report that physicians generate diagnostic hypotheses early in the clinical encounter and that these may be either specific or general. But, again, no definition of these terms is given.

The process of defining categories and indicators for content analysis has yielded a definition and examples of two forms of interpretative response to clinical data which are evident in the present accounts of students, house officers and registrars. These two forms of response may be analogous to the general and specific hypotheses discussed above, yet closer examination will indicate that 'hypotheses' may not be the most productive term to apply to them. It has already been suggested that 'hypothesis generation and testing' is a needlessly limiting conceptualisation of the diagnostic thinking process (see Chapter Three). The categories of this study refer to 'pre-diagnostic' and 'diagnostic' interpretations of clinical information.

A pre-diagnostic interpretation could not be described as a 'diagnostic hypothesis', not even as a 'broad or general diagnostic hypothesis'. Instead, it appears to represent the subject's clearly interim, working and flexible identification of the general location of the piece or array of information in some segment or section of his own cognitive structure of knowledge and experience. In other words, it may be seen as a working interpretation, not an hypothesised diagnosis. It is the beginning of the process of trying to make sense of the information elicited by referring to other information stored

in memory. This can be done in general terms by locating broad areas of relevance which could not be a diagnosis, or in specific terms by locating a precise, unambiguous meaning which would be acceptable as a diagnosis. Thus the interpretation refers to an extrapolated context or array of information of which the information already elicited could be an exemplar if accompanied by certain other items of information.

Considering the first part of Hypothesis 10 which refers to multiple interpretations in all groups of subjects, this is substantiated as Table 10.2 indicates. Students, house officers and registrars each reported approximately seven pre-diagnostic interpretations of the information and four or five diagnostic interpretations. Reference to Appendix 30 shows that the range across all groups of pre-diagnostic interpretations is from zero to 18, and for diagnostic interpretations from zero to 11. It is certainly the case, then, that multiple interpretations at both levels of specificity are made by all groups. This reflects the findings of Elstein et al (1978) and the McMaster group (see 3.1.2) who also point out that students and physicians do not differ in the number of hypotheses they generate. Our findings substantiate this conclusion and add house officers to the list. All groups appear to have equal interpretative capacity.

The Michigan group find that the average number of hypotheses held at any one time ranges between four and seven, while the McMaster group identify a range of three to nine with an average of 5.5. Table 10.2 shows that our subjects made an average of approximately 11 or 12 interpretations during the course of the clinical interview. It has already been suggested (3.1 above) that both the Michigan and McMaster groups have tended to overinterpret their data and attribute to it an accuracy, completeness and rigour which is unjustified. In the present instance this tendency, together with the inclination to quantify which the formulation of hypothesis generation and testing has encouraged, has yielded a misleading

and limiting view of the diagnostic thinking process.

Let us first consider what it might mean when physicians or students are described as holding a certain number of hypotheses 'at the same time', disregarding for the moment our contention that account gathering methods are not sufficiently accurate to allow statements of such specificity. It must be presumed that the hypotheses under discussion are seen as being 'held' in short term memory as though they were, in a sense, ideas to remember and test out in the real world by establishing the presence or absence of their concomitants. This is surely a description which lacks the dynamism of the actual process, and fails to consider the context of antecedent and subsequent thinking. It is only a partial argument, dealing with a part of the diagnostic thinking process which has become generalised and expanded to usurp the description and definition of an entire process. We have evidence of multiple hypotheses or, as we prefer, interpretations, but instead of stopping our analysis with that finding, we must ask how such a cognitive phenomenon can arise. What happens as a result of it is discussed in relation to later research hypotheses.

In essence, for a number of different interpretations to be made of the same information array, or of subsections of that array, or intersections of that array as it gradually increases or changes in certain ways, it is necessary for the interpreter to have as many different perceptions of the information array as he has interpretations. Different perceptions, logically, can arise in two ways. Firstly, the same items of information, where they constitute a complete array, can be rearranged cognitively in some manner to have different inter-relationships, roles, relative dominance and so on (see 4.1.2 above for the discussion of the forceful feature). In other words, the array can itself be structured in different ways by the interpreter, so that it has different meanings for him. Secondly, the array of information can be rearranged cognitively in the same way as described but in

an extrapolated context of information which might be present but, as yet, is not proved to be so. In other words, the information can be given structure and meaning by fitting it in to or relating it to a wider set of information already structured and stored by the interpreter. We may see here a parallel with Peel's (1971) describer and explainer thinking. The multiplicity of pre-diagnostic and diagnostic interpretations suggests that information as it is elicited from the patient can be and is structured as part of many possible extrapolated contexts of learned, memorised or otherwise stored information.

It is clear from the discussion so far that Elstein et al's (1978) differentiation between cue interpretation and hypothesis generation is not considered tenable. Instead, it is suggested that these constitute two aspects of the same process.

Given the derivation of interpretations (or hypotheses) as suggested, it is clearly inappropriate to contend that a certain number of these may be held at any one time. A certain number may be dealt with at any one time, but that number is probably as dependent upon the clinical information itself and its relationship to the various potential interpretations, as upon the needs of the student's or clinicians own thinking processes. This echoes a point already made (3.6 above) that multiple hypothesis formation is less likely to be for reasons of economy of cognition than a simple function of the multiplicity of potential extrapolated contexts in which the given information can be embedded.

Having considered the possible cognitive processes antecedent to the hypotheses or interpretations, we may consider the subsequent processes. Neither the Michigan nor McMaster studies provides information concerning the fate of hypotheses once held but now discarded. It is shown later in this chapter that the capacity of the interpreter to structure information in inappropriate ways is not always severely limited by either fact or logic. The accounts also demonstrate that interpretations may wax and wane in relative importance and currency

or may be not mentioned between the beginning of the account and the review of the case at the end. Yet during this time, all interpretations are potentially current and present, or, rather the interpreter has the ability to structure the information in many ways other than those he is currently discussing and evaluating. It is this potential which is surely the important aspect of the diagnostic thinking process. The ill defined problem potentially can become well defined in many different ways which have varying degrees of congruence with reality. This provides another strong reason for rejecting the hypothesis counting approach towards describing diagnostic thinking. Such an approach obscures the dynamism and creativity inherent in the process.

The final point to be made in relation to Hypothesis 10 concerns its second part, which suggests that no differences will be observed between students, house officers and registrars in their relative use of pre-diagnostic and diagnostic interpretations of the clinical information elicited. The non-statistically significant interaction term of Table 10.1 suggests that this hypothesis is accurate in its assumption. The value of F for columns (categories), adds to this finding by showing that the incidence of pre-diagnostic interpretations is statistically significantly more frequent than the incidence of diagnostic interpretations. We may consider reasons for this finding.

Firstly, it is not the case that pre-diagnostic and diagnostic interpretations appear during the clinical interview or diagnostic thinking process in any particular order. That is to say that the diagnostic thinking process cannot be characterised as being one of gradual specification and definition. The making of either type of interpretation primarily must be dependent upon the information elicited. If that information is specifically diagnostic of one disease, the clinician will not think pre-diagnostically simply because the interview has only just commenced. It is possible that in some circumstances he would be well advised to do so, but that is not the point under discussion.

Secondly, given that information is elicited linearly, frequently needs clarification, and that the clinician is constantly responding to the flow of information, it would seem reasonable that he could not be specific as often as he could make interim judgments. In terms of structuring the information, or relating it to stored, structured information, it may be that the structures of information which can be labelled with the name of a diagnosis are of a different order from the structures of information which can be labelled with a pre-diagnostic interpretation. It is possible that the content may or may not have features in common. These are all matters of pure speculation and would require further investigation and research. We may only state at present that relative use of pre-diagnostic and diagnostic interpretations does not differ across the groups of subjects of this study who all favour the former over the latter. It is also worth noting that a pre-diagnostic interpretation can be just as correct or just as incorrect as a diagnostic one in relation to the patient's actual diagnosis.

11.3 Hypothesis 11

"(Elstein et al (1978) find that, in terms of time of onset, the first diagnostic hypothesis is generated by ten per cent of the way through or within the first five minutes of the clinical interview. Barrows et al (1978) find that the first diagnostic hypothesis is advanced, on average, within 28 seconds of the appearance of the main complaint).

The present study will show that students, house officers and registrars make immediate interpretative or evaluative response to initial items of clinical information received whether or not these constitute the patient's main complaint."

This discussion concerns subjects' responses to information elicited during the first replay period which varied in mean length from 41 to 49 seconds and actually ranged from 13 to 82 seconds. For reasons explained in section 10.2.2 above, revelation or not of the patient's main complaint is not considered a relevant factor.

The intended aims of Hypothesis 11 are threefold. Firstly, it might provide a basis for informed discussion of theories of

pattern recognition as an element in the diagnostic thinking process. Secondly, it might reinforce previous criticisms of the theory of hypothesis generation and testing as inadequate, misleading and limiting because of its tendency to divert attention to one aspect and section of the process only and generalise those findings to other parts of the process not seriously analysed. Finally, it might deepen our own interpretation of the thinking process under consideration.

Although theories of the diagnostic thinking process as pattern recognition have been discussed and rejected (3.2 above) they must again be considered in their role of explaining the student's or clinician's cognitive activity prior to generation of the first hypothesis. Barrows et al (1978) and Scadding (1967) suggest that generation of the first hypothesis (although, apparently, not others) occurs as a result of pattern recognition. We still reject this interpretation, but must forward an alternative one if the criticisms made are to be other than destructive.

It will be remembered that the content analysis categories of Hypothesis 11 are as for Hypothesis 10 (pre-diagnostic and diagnostic interpretations). In addition, however, is one concerning the subject's judgment of the need for further clarifying enquiry not stemming from either pre-diagnostic or diagnostic interpretations (see 10.1.2 above). Results indicate that only two students and one house officer failed to make a response which could be classified into one of these categories. Of these, one student had experienced difficulty in hearing what the patient had said, the other had discussed the patient's holiday location, and the house officer had been informed by the patient that she had chronic lymphatic leukaemia and he only commented during the first stop that he had thought it was a pity that she had told him that. These three subjects, then, are exceptional for particular reasons and cannot be taken as invalidating the following discussion in any way.

It was found that almost two thirds of the subjects made a response in one of the categories while almost one third made

responses in two categories and one student actually made responses in all three categories (Table 10.6). Pre-diagnostic interpretations were the most common response (almost half of all responses were in this category) while diagnostic interpretations and a judgment of need for further clarifying enquiry were almost equally frequent in the remaining half of responses made. As discussed in 11.2 above, we cannot draw any conclusions about the subjects' thinking processes from this relative distribution of responses across categories because the effects of information content cannot be determined. This point is also made in section 10.2.2 itself, where attention is drawn to the major lesson of these findings which is that subjects can and do make active, interpretative or evaluative response to clinical information as soon as it is elicited. As Hypothesis 11 suggests, this is so for all groups of subjects. In addition, Table 10.8 shows that the relative distribution of responses across categories (a), (b) and (c) is not statistically significantly different according to status of the respondent. Elstein et al (1978) suggest that hypotheses are generated by associations from clusters of a few cues. We would suggest that even a single piece of information can be structured in an extrapolated context and interpreted.

Although the point made is a small one, it is of considerable importance. We have shown that if clinical information is available, the student and the clinician make active response to it. This is substantiated in the Piagetian viewpoint "that human beings are not blank slates which passively receive the world; rather that they actively structure it" (Case, 1973). In 77 per cent of instances (64 out of 83 responses) the response is in the form of an interpretation. In other words, the subject has located an appropriate (or inappropriate, but nonetheless possible) extrapolated context of information and has cognitively structured (or interpreted) the information elicited accordingly. We would therefore reject with even greater certainty theories of pattern recognition which imply the passive reception of information until it is recognised as matching some other array. The unlikelihood of this as either a logical or cognitive phenomenon has already been dis-

cussed in section 3.2 above. We now have presented evidence to support that theoretical argument. In principle, arguments of pattern recognition can adequately be refuted by the assertion that the human brain is not a passive receptor of information but is, even at its least complex level, an active assimilator, or, as Neisser (1967) asserts, "seeing, hearing and remembering are all acts of construction, which make more or less use of stimulus information depending upon circumstances". So it must be with pattern recognition as proposed in the field of theories of the diagnostic thinking process. Reminding us of a previous discussion, Neisser (1967) points out that: "Ill defined categories are the rule, not the exception, in daily life". This being so, he reminds us also of Bruner's emphasis "that pattern recognition depends on the identification of specific features or attributes of the stimulus". This must surely be the case if, as according to Barrows and Scadding, a process of pattern recognition leads to an hypothesis rather than an identification! The contention seems to be a logical, and psychological, fallacy; for if it is such a process of identification of specific features and does, which is more important, lead to an hypothesis rather than an identification, then surely the process is one of interpretation by structuring. It would seem that again, perhaps, concentration on the formulation of hypothesis generation and testing has precluded useful analysis, directing attention to the search for the first hypothesis and not to explaining how that hypothesis is derived.

The process of content analysis has revealed three possibilities in relation to the derivation of meanings. Two of these are the formation of pre-diagnostic and diagnostic interpretations of information, respectively, which we suggest is by complementary cognitive operations of structuring and extrapolation to other structured contexts. However, the third category requires some discussion since its congruence with our overall interpretation of the diagnostic thinking process is not immediately clear. Section 10.1.2 names this third category, called category (c), as "Judgment of the need for further general or clarifying enquiry, not stemming from either pre-

diagnostic or diagnostic interpretations". It is defined as: "Where S enquires further about the patient's symptoms, signs, etc. for clarification OR where S seeks to clarify the patient's statement NOT where S is seeking a particular piece of information based on his own expectations". An indicator of this category may be "I was asking how the pain affected him" or "I asked her how she didn't feel well". Table 10.5 shows that 23 per cent (19 out of 83) of all responses were in this category but that it was the least frequently used of all response categories identified.

Hypothesis 11 suggests that all subjects make either (or both) interpretative or evaluative response to initial items of information elicited. Category (c) clearly represents an evaluative response which reflects the subject's inability to structure or interpret the information to his own satisfaction without further clarification of the patient's meaning. It is important to note that we say "to his own satisfaction". It would be untenable to suggest that a subject, if presented with the information that the patient "didn't feel well" or "had a pain", could not extrapolate from that to many contexts, both pre-diagnostic and diagnostic. But to do so would be grossly inefficient. It would seem, therefore, that the subject, perhaps, makes a judgment about how well defined the problem should become before what the Michigan and McMaster groups call 'hypothesis testing' can usefully begin. By 'well defined' in this context, we must think in terms of structures or extrapolated contexts which the subject can identify with sufficient subjective certainty or from which he can work effectively. Identification of category 11(c) concerning the need for further enquiry suggests some evaluation of the interpretative process itself or of its initial outcome and a greater or less willingness or ability to work within one structure rather than another (see 11.12 for further discussion of this point). No grounds or evidence are available which indicate whether or not this is a person or personality dependent phenomenon, or whether all subjects would respond to information as vague as "didn't feel well" or "had a pain", in the same way.

Ultimately, as Peel (1971) indicates "the decision turns on what the judge wants to fulfil". The question, here, is whether or not every judge wants to fulfil the same objective in the same way.

In conclusion, Hypothesis 11 has given rise to further consideration of the process of interpreting clinical information and has thrown some light on the conditions of that process. In addition, the previous rejection of pattern recognition in the diagnostic thinking process which was on theoretical grounds, has received empirical support.

11.4 Hypothesis 12

"Accounts given by students, house officers and registrars will indicate that the diagnostic thinking process involves working within a cognitive context or contexts extrapolated from the clinical information available."

Results and discussion relevant to Hypothesis 12 should clarify the meaning of "interpretation" as it has been used in categories 10 and 11(a) and (b), which refer to pre-diagnostic and diagnostic interpretations. These two categories are also used for content analysis relevant to Hypothesis 12 but are augmented by a third category identified as "expecting, searching for or planning to search for specific features (symptoms, signs, tests, etc.) of disease or treatment of disease". This category excludes enquiry which forms part of the routine or systematic review. Its close relationship to the concept of hypothesis testing as promulgated by the Michigan and McMaster groups should be quite obvious. Indeed, the results and discussion of Hypothesis 12 will substantiate the findings of those workers.

Much of the potential discussion has already been presented in relation to Hypotheses 10 and 11. The concept of structuring and its relationship to the extrapolated context has already been explored. Table 10.9 shows that in only two of 198 instances is no response made in any of the three categories of Hypothesis 12, but that all subjects made response in at

least two of these categories and 64 out of the 66 made responses in all three categories. The present findings, then, surely reflect those of the Michigan and McMaster studies which those groups characterise as hypothesis testing. It is intended, here, that category 12(c) which concerns expecting or searching for specific items of clinical information, will enable us to be more specific about the nature of this cognitive process.

De Groot (1965) refers to "the omnipresence of anticipations" and states that, in chess playing, "practically every operation in the thought process serves to find out whether things specifically anticipated are or are not confirmed". De Groot, like the Michigan and McMaster groups, conceptualises his subjects' behaviour as hypothesis generation and testing. Elstein et al (1978) use this same framework precisely, stating that clinical data are evaluated in terms of their fit to anticipated findings. They add that each cue is designated as positive, non-contributory or negative with respect to a particular hypothesis. However, this conclusion would appear to be a paramorphic one, and an artefact of their research methods rather than actually reflecting their subjects' thinking processes.

The present study strongly reinforces the explanatory concept of searching for anticipated findings as the 100 per cent frequency of response in category 12(c) shows. As an example of a subject's account of this type of thinking, we may quote from a house officer:

"Well, with that kind of history I was immediately thinking of myocardial infarction. I was trying to keep an open mind and at the same time thinking: I must go through these first. I must go through the typical things for an infarct". (H.O. 04)

In other words, having located a possible structure, the subject needs to define whether or not the 'fit' which he has achieved extends to the rest of the information. If so, then a diagnosis is made. If not, then the information elicited must be compared with other structures in the same way until one is found that does fit. We may quote from the account of a student, talking about the same aspect of his thinking

processes:

"You always expect other symptoms. If he's got a cardiac complaint like that, especially coming on for no apparent reason, you have to think of pain or anything else or whether he felt dizzy, fainted, nausea, or anything which may be associated with his chest complaint, and, in fact, I found out that he did have dizzy spells, so I pursued that a bit further.

(And what were you thinking at the time?)

Well, trying to work out the differential diagnosis of the various heart conditions that he might have".

(S. 20)

This account is interesting in that it hints at a more precise meaning of both structuring and interpreting. In this case, the subject is anticipating some subset of a possible range of concomitants. Perhaps we may perceive here, a point of differentiation between pre-diagnostic and diagnostic interpretations. S.20's pre-diagnostic interpretation 'cardiac complaint' clearly subsumes a large range of symptoms, signs and other clinical information. When S.20 has established the presence or absence of some or all of these, then he may have worked out the differential diagnosis (or diagnostic interpretations) he was trying to achieve. In other words, more information of a precise type might enable him to designate a number of possible structures of the information elicited and eventually determine which of the extrapolated contexts is the most likely one.

This, then, would appear to be a more precise, cognitive, description of the process of hypothesis testing when seen in the context of structuring and interpretative activity. It remains an interesting point, however, to speculate upon how much or what types of information give rise to pre-diagnostic and diagnostic interpretations, respectively; what types of information are stored in these structures; whether they are the same or different and similarly or differently organised; and whether there are individual differences in these features across subjects within and between groups. The present data can throw no light upon these dark areas.

11.5 Hypotheses 13 and 17

" It will be observed for students, house officers and registrars that re-interpretation of clinical information occurs during the course of the clinical interview due to:

- (a) New thoughts occurring about already interpreted clinical information when no new information has been added.
- (b) New clinical information being elicited to facilitate re-interpretation of clinical information already elicited and interpreted."

" Results will not indicate the mechanism of interpretation and re-interpretation of the array of clinical information as it is accumulated throughout the clinical interview."

Hypotheses 13 and 17 are complementary and so are discussed together. In sections 3.1.1 and 3.1.2 above are defined questions which the Michigan and McMaster studies did not appear to address. Hypotheses 13 and 17 attempt to make good some of these omissions by initiating discussion of the following questions: How are cues cognitively manipulated ? What process of structuring generates hypotheses ? How do new hypotheses arise ? How are relationships between cues allocated and re-allocated ? The categories of Hypothesis 13 which should facilitate discussion of these questions are identified as: "Re-interpretation of clinical information, when no new information has been added". This is category (a) and it reflects part (a) of Hypothesis 13. Category (b), reflecting part (b) of Hypothesis 13 is identified as: "Re-interpretation of clinical information arising from the addition of new information". For both categories, we are considering only the restructuring or re-interpretation of information which has already been structured or interpreted in some way by the subject. Thus, original or primary interpretations of information are not the subject of content analysis. This is because the data are not sufficiently rigorous to enable us to say precisely at what point an interpretation actually arose. For example, during the first replay period H.O. 01 elicited (i) that the patient had experienced pains

across her chest and stomach area; (ii) that these had led to sickness with meals; (iii) that she vomits mucus during the second course of the meal; (iv) that she had suffered from this complaint for four to five months. H.O. Ol interpreted this as 'a gastrointestinal problem'. However, from these data, we cannot identify at what point in the actual interview this interpretation was first made by H.O. Ol. We only can say that it was made at some point during the flow of information which presented these four points. Thus, the original stimulus or stimuli to structuring or interpretative response cannot be identified precisely. However, changes in structure or interpretation can be identified since subjects invariably pinpoint the cause of their change in thinking. Categories 13(a) and (b) therefore concern restructuring and re-interpretation only. It follows from this that if a subject does not display responses in categories 13(a) or (b), we may infer that new information elicited either gave rise to entirely new interpretations or was relevant to interpretations already made and was structurally congruent with the structural frameworks selected by the subject already. In this case, new information would not cause the subject to re-interpret. Thus, if frequencies of response in categories 13(a) and (b) are low, we can justifiably infer that re-interpretation or restructuring of information is not a cognitive skill often used and that the alternative skill of selecting among primary interpretations is more in evidence. The effectiveness of this would depend upon the accuracy and appropriateness of primary interpretations.

Table 10.10 shows that 39 per cent of subjects did not make any re-interpretations of information during the clinical interview. Of the remaining 61 per cent of subjects, 58 per cent made re-interpretations on the basis of new information being added to the array and 20 per cent made re-interpretations in the absence of new information. Of these, 17 per cent made both types of re-interpretation. The non-statistically significant value of the Kruskal-Wallis H indicates that re-interpretative response patterns do not differ across groups of

subjects. We therefore find Hypothesis 13 substantiated. Subsequent discussion indicates that Hypothesis 17, as hoped, is not substantiated in that the data provide useful information about mechanisms of interpretation and re-interpretation of clinical information.

Before considering the nature of re-interpretative responses, let us first consider the relative frequency of re-interpretation and lack of re-interpretation, and the implications of this for our understanding of the diagnostic thinking process. It is worth reiterating that there is no intrinsic merit in either form of cognitive process. We could suggest that re-interpretative behaviour indicates a flexibility of thought, lack of tendency to rapid closure and field independence. These qualities would usually be considered desirable. However, this judgment cannot be made in the present circumstance since it may be the case for the 39 per cent of subjects who did not display re-interpretative responses that their primary interpretations were sufficiently accurate not to warrant it, or that the clinical information was elicited in an order and manner that located it easily within the appropriate stored structure. On the other hand, in some cases failure to re-interpret could cause diagnostic errors to be made (see 11.9 below). The same types of argument, in inverse form, can be applied to the 61 per cent of subjects who displayed re-interpretative responses.

In summary of this argument, we cannot state whether or not the categories of thinking here under consideration are, in absolute terms, desirable or not, or appropriate or not. Such judgments can only be made in terms of individual diagnostic problems and thinking processes. Either type of thinking (re-interpretative and non re-interpretative) could be inappropriate or appropriate depending upon the circumstances of the case. However, it can be said that both types of thinking occur in the solution of diagnostic problems, during the clinical interview. It can, therefore, also be concluded that the diagnostic problem solver should have both types of response in his

repertoire of potential cognitive processes to be applied to the diagnostic problem. It can also be suggested that the diagnostic problem solver should be aware of his capacity to think in these ways and his manner of application of such thinking processes.

Let us now consider the two types of re-interpretative behaviour identified. Re-interpretation in the absence of new information is less common in the samples of subjects than is re-interpretation in the presence of new information. Nonetheless, 20 per cent of subjects made responses in the former category. These are spread across the three groups of subjects relatively equally (three students, six house officers, four registrars), therefore explanations based on experience, knowledge, skill, etc. would seem untenable. All we can infer with certainty is that, for some reason, the subject did not extrapolate from the information to all possible contexts at the time of initial presentation, but made new extrapolations later in the process. From the indicators (10.1.4 above) it can be seen that either the new interpretations seemed to occur quite suddenly, for no apparent reason, or they occurred as a result of a conscious effort to find new possible explanations of the given information. In both cases we must suppose that the problem solver is in the process of manipulating and working with the information, such that new structures or interpretative contexts are located. Most subjects gave no explanation of the phenomenon, but merely reported it:

"I think, probably, I thought about PBC when she was a middle aged woman who complained of generalised itching ... I thought of Hodgkin's when I was, sort of in the odd moments, in the gaps in the session. Sort of racking my brains for other causes of generalised itching". (H.O. 02)

"While I was going through the routine stuff, it came into the back of my mind that she might have some polyuria or polydipsia associated with the basal condition". (H.O. 05)

It may well be that the trigger to H.O. 05's restructuring or re-interpretation was his reaching the part of the routine

history that made him ask about the patient's genito-urinary system. Thus the stimulus to restructuring was not new clinical information, but a new area of enquiry and so a possible new context for seeking relevant interpretations. Uses of the routine history are discussed below (11.7) but stimulation to new perceptions of the information and new contexts for interpretation may be a useful spin-off.

Having considered re-interpretation in the absence of new information to the extent that the data will allow, let us consider the nature of re-interpretation in the presence of new information. Figures 11.1, 11.2, 11.3 and 11.4 provide schematic representations of four instances of re-interpretation of information. The caption for each provides a description of the process of restructuring involved. Figure 11.1

<u>Information Elicited</u>	<u>Subject's Response</u>
Shaking Diarrhoea	Infective cause
Gastrectomy	" "
SOB with meals Stomach operation Sweating Palpitations	Dumping syndrome
Diarrhoea	Side effect of gastrectomy
<u>Figure 11.1 Example of Re-interpretation (Cognitive Re- structuring) of a Symptom (Diarrhoea) in the Presence of New Information. Category: 13(b) Subject: S. 02</u>	

shows a simple instance of new information giving a completely new context and facilitating extrapolation to a context which, in the first instance was not considered. The subject's initial response (infective cause of diarrhoea) is tied very closely to the information given and does not separate the two pieces of information to allow very wide interpretation. However,

<u>Information Elicited</u>	<u>Subject's Response</u>
Nose-bleed	Nasal tumour/naso-pharyngeal/ tumour/nose-picking/hypertension/ uraemia/nephritis.
High blood pressure	Nose-bleed due to hypertension.
Inhaler	Appearance caused by steroids.
Cushingoid appearance	Steroids for asthma. Steroids
Asthma	affect immune response, therefore
Steroid treatment	infection. Steroids affect blood
Infection	pressure, therefore nose-bleed.
	Problem: to adjust steroid regime.

Figure 11.2 Example of Cognitive Restructuring of a Symptom
(Nose-bleed) by Embedding in a New Context. With Complex Re-
structuring of an Information Cluster. Category: 13(b)
Subject: S.O. 4

However, when a new, and crucial, piece of information is presented (gastrectomy) the subject does not restructure his interpretation. This occurs only after more new information is elicited which itself gives rise to a new context.

Figure 11.2 shows an instance of restructuring complicated by apparent confirmation of one possible interpretation and a refinement of this by means of a complex disentangling, re-ordering and restructuring of an array of information which yields a further precise context and a problem solution. Figure 11.3 presents a relatively less complex instance of restructuring where the information seems gradually to lead the subject away from an incorrect interpretation and towards the appropriate context. Figure 11.4 presents an instance in which the subject was unable to extrapolate to the correct context from the information given and even characterised it as "outside my experience" until a forceful feature⁽¹⁾ which was also the cue⁽²⁾ ("pacemaker") allowed him to identify the appropriate extrapolated context.

Footnotes (1) and (2) - see end of chapter

<u>Information Elicited</u>	<u>Subject's Reponse</u>
Collapsed lung Cardiac lung	Need to elaborate
Collapsed two years ago	Heart attack/stroke
Bronchitis prior to collapse	Fallen/breathing trouble/heart trouble
No chest pain Breathing trouble	Acute attack of something res- piratory, not cardiac/acute chronic respiratory disease.
<u>Figure 11.3 Example of Cognitive Restructuring From Cardiac to Respiratory Context. Category: 13(b) Subject: S.12</u>	

<u>Information Elicited</u>	<u>Subject's Response</u>
Aged 82 Three blackouts	Atherosclerosis/transient ischaemic attacks/vertebro-bas- ilar drop attacks/minor CVAs
First blackout lasted four hours	Not transient ischaemic attack/ not drop attack
After each attack:	
No confusion	Not recognised
No weakness	"Outside my experience"
Normal immediately	
No other illnesses	
Good general health	" "
No palpitations	" "
Active	
Climbs stairs	
Pacemaker	Stokes Adams attacks/heart block/ after silent infarct.
<u>Figure 11.4 Example of Initial Failure to Find the Appropriate Extrapolated Context, Followed by Rejection of Inappropriate Interpretations and Final Restructuring After New Information Facilitates Correct Interpretation. Diagnosis: Stokes Adams Attacks due to Ischaemic Heart Disease. Category: (13)b Subject: S.19</u>	

From these four examples alone we see a wide range of structuring and restructuring mechanisms operating. The subjective experience was described in one account as follows:

"Cough and shortness of breath immediately tells you which part of the body you're dealing with. And then the cogs start going in your head, asking what causes cough and shortness of breath. Is it chronic bronchitis or is it something else?". (H.O 16)

But 'the cogs start going' covers an apparent multitude of processes. For example, a single item of information alone can cause a re-interpretation of information, or it may take multiple new items of information before an array becomes amenable to re-interpretation (as in Figure 11.1). Sometimes the very interpretation itself gives rise to other new interpretations of the same information. For example:

"I wondered at that time if he'd had superior vena cava obstruction as well. There was a lot of swelling in the arm. That was just something that came flying through. When you think of carcinoma of the lung you wonder whether it might have some exciting non-metastatic complication". (S. 17)

Occasionally, two previously separate interpretations become associated or linked to form one, integrated superordinate interpretation. For example, subject R. 12 had pursued liver disease and renal failure as two separate phenomena until it "struck" him that in the particular patient they may both be part of the same disease (Weil's disease). Likewise subject S. 05 had interpreted a stroke, shortness of breath and fits as three separate problems until he extrapolated to the context of arrhythmia and concluded that:

"I suppose it's possible that an arrhythmia threw off an embolus that went to the brain. It might not have been an artery snapping, it might have been a thrombotic or embolic episode". (S. 05)

Thus, three separate interpretations became two: arrhythmia with cerebral embolus and epilepsy.

It is possible that each diagnostic problem solving process differs from every other in its details of mechanism and thinking. It therefore would seem unwarranted to characterise

individual cases further, except, perhaps to indicate that different types of information may be differently amenable to interpretation, as the identification of pre-diagnostic and diagnostic interpretations may cause us to believe. One subject's account illustrates this phenomenon quite vividly:

"Yes, talking about the weakness. Can't get away from it, can I? It's the most difficult, vertigo and weakness. If a patient comes in and says "I feel dizzy and weak", it's the syndrome that makes an internist go "Ugh!" and cringe and crumble. Then he let me into this gem that he had discomfort in his belly ... and I elaborate on this". (R. 02)

"Discomfort in his belly" was information from which R. 02 could easily extrapolate; "weakness" was not.

In summary, we may consider what general principles can be inferred from our discussion of results relevant to Hypotheses 13 and 17 which concern the cognitive processes of interpretation and re-interpretation of the changing array of clinical information. It has been shown that re-interpretation may occur either in the presence or absence of new information being added to the array, and it is suggested that in each case re-interpretation occurs as a result of cognitive restructuring of information and consequent embedding of that information in a new extrapolated context of possibly related information. It is suggested that re-interpretation and lack of re-interpretation may each be either appropriate or inappropriate depending upon the particular diagnostic problem in relation to which it occurs. Consequently, the diagnostic problem solver must be aware of these cognitive processes, have them in his repertoire of thinking processes and be able to monitor and apply them appropriately.

No differences are found between the three groups of subjects in incidence of re-interpretative responses and so these cannot be related to other variables such as knowledge, clinical experience and so on. However, this does not preclude the content of these processes varying in appropriateness from group to group, while the structure of the processes themselves is common to all groups. The nature of stimuli to interpre-

tation is considered and found to be heterogeneous. It is concluded that the vast range of clinical problems, presentation of those problems, and potential interpretations precludes useful classification of cognitive processes or mechanisms from the point of view of content. Instead, it is suggested that the general cognitive activity of interpretation and re-interpretation with its difficulties and facilities and its relationship to the stored, structured knowledge of the subject, must be part of the conscious and serviceable knowledge of the diagnostic problem solver. From the point of view of teaching and learning, the relationship between the structure of information as it is elicited from the patient and responded to by the clinician is an important aspect of this special knowledge. Why should some types of information cause the clinician to respond 'Ugh!' and 'cringe and crumble' when other types of information give rise to multiple interpretations and a clear line of enquiry? Such questions as this reflect upon the appropriateness and usefulness of the structure of stored information and upon the subject's means of gaining access to that store. Awareness of such problems is a necessary prerequisite for their solution.

11.6 Hypothesis 14

"No differences will be observed between students, house officers and registrars in use of strategies for selection between competing interpretations of clinical information."

We have so far considered the factors antecedent to the interpretative response of the diagnostic problem solver. Hypothesis 14 concerns the eventual fate of those interpretations. We are, therefore, considering an activity analogous to that of Elstein et al's (1978) 'hypothesis evaluation' stage.

The content analysis categories of Hypothesis 14 (see 10.1.5 above) are identified as 'Active confirmation of an interpretation' (category (a)) and 'Active elimination of an interpretation' (category (b)). The logically complementary category of 'passive elimination' is omitted since the data do not

allow determination of whether acceptance of one interpretation actually implies rejection of another as its corollary. The final category is identified as "Postponement of either confirmation or elimination of a possible interpretation with or without stated differential likelihoods". It is noted that these categories cannot be seen as absolute. Categories (a) and (b) may be subject to reversal given new information. Thus the categories describe cognitive operations which may be present during the diagnostic thinking process. It is considered psychologically untenable to suggest that an interpretation, once made, can be totally expunged or confirmed without chance of later doubt or is not capable, when eliminated, of being later re-identified. We are interested only to establish the presence or not of the thinking processes represented by categories 14(a), (b) and (c) in each group of subjects.

The results reported in 10.2.5 above show that the three groups of subjects do not differ in their frequency of use of each of the categories of response identified for Hypothesis 14. Neither do they differ in their use of multiple or combinations of categories (Table 10.11), although we find that there is a statistically significant difference across all groups, in use of either one, two or three categories. The use of two categories is more common than the use of either one or three, although no particular combination of categories seems to predominate. It is particularly interesting, in the light of the discussion of section 3.3 of the predominance of tendencies to verification rather than refutation and to use confirming rather than refuting evidence, that there is no tendency for category (a) responses to predominate. These findings may be attributed to a number of possible causes, but not to any variables which differentiate between groups. They cannot be attributed, therefore, to factors such as knowledge, skill or clinical experience. Of course, the outcome or content of the thinking processes under consideration may be correlated with such variables as these, but not the processes themselves. It may be reasonable to suggest that differences in strategy for selection among competing interpretations of information are dependent, to some degree, upon individual differences in personality or cognitive

style. For example, the nature of acceptable 'proof' may vary from subject to subject. Alternatively, the finding that subjects tend to apply two selection strategies more often than either one or three may be due to logical characteristics of the problem itself. When multiple interpretations are made of the same information or of overlapping sets of information, then these interpretations cannot be independent. Thus a decision about one (for example, actively to eliminate it) must have repercussions for the others (they may become stronger and more prone to deferred decision or active confirmation). However, the discussion must remain at a level of speculation or theory only, since the data do not provide information concerning the fate of each interpretation made by each subject. Indeed, it is doubtful that such information is attainable.

The final possibility to be considered must be that the strategy used by a subject for selection between competing interpretations of clinical information is dependent, to some extent, upon that information itself. There are cases in which important information is not available or is unclear and so, unless extremely reckless, the student or clinician must defer judgment. There are cases in which the symptoms and signs (for example, a pacemaker, a recognised syndrome, or a special treatment) cannot be mistaken as indicative of more than one condition and are sufficient to diagnose that condition with the maximum degree of certainty possible.

We must conclude, therefore, that an array of factors may be operating to determine any subject's strategy for selection among competing interpretations of clinical information. These factors may derive from the characteristics of the problem solver or the content or structure of the problem itself, or the logical relationship between competing interpretations. These factors may vary in relative prominence and importance from case to case and subject to subject. For reasons such as this Elstein et al's (1978) identification of rules for diagnostic decision making based on the numbers of positive and negative cues must be rejected, not primarily for their probable artefactual nature, but for their implicit simplistic model of the variables inherent in the diagnostic problem and process.

Likewise, Barrows et al's (1978) finding that any piece of clinical information elicited by his physician subjects was relevant to about three hypotheses, ignores the logical characteristics of the diagnostic problem and process. This point has already been discussed in 3.1.2 above.

Elstein et al (1978) contend that clinical data are evaluated in terms of their fit to anticipated findings. Our discussion of the nature of interpretation and structuring would reinforce this viewpoint. Subjects are prone to discuss their thinking processes in terms of establishing 'fit'. For example:

"By that stage it was getting very unsatisfactory because, firstly, I presumed, as I had done from the start, which is a bad thing, that the big episode was an infarct. Yet everything I've elicited about it, there's something unsatisfactory about. It doesn't quite fit in with that. Second, the giddiness attacks which I automatically presumed were some kind of angina. But they don't seem to fit. And they weren't vertigo, so they involve something else. So I haven't tied the two things together and I haven't really made a satisfactory diagnosis. And then I'm getting a history of diabetes which would also fit in ischaemic heart disease. But it's not really adding up." (H.O. 04)

Elstein et al (1978) also suggest that cues are designated as positive, negative or non-contributory with respect to any particular diagnostic hypothesis. Although this may be overstating or over formalising the precision of the diagnostic problem solver's thinking, it would seem likely that some similar, but less conscious, process occurs. For example:

"Well, I was thinking of general things that could be wrong with his heart. I mean, he hadn't had a coronary almost certainly, because the symptoms weren't right. He hadn't had chronic angina - the symptoms definitely weren't right for that. It didn't sound as if he was having any arrhythmias. Well, he did actually mention palpitations, but they were very regular palpitations, just as if the heart was beating hard and normal rather than irregularly. I was coming on to he must have a valvular defect and I was going to come on to his past medical history to see if anything was in that which I was expecting to find, really". (S. 20)

If S. 20 had been asked to rate each piece of clinical information as positive, negative or non-contributory, he could undoubtedly do so. However, such an exercise would not demonstrate that his usual thinking process involved such a tactic. It must also be noted that, as with the palpitations in S. 20's case, the information has to be evaluated in its own right before it can be judged in relation to any wider interpretations of the patient's problems. The mechanisms of confirmation, elimination and deferred judgment of interpretations are illuminated further below in relation to ideas of psychological probability (see 11.11).

In summary, we have shown that competing interpretations of information are dealt with in at least three different ways; by confirmation, elimination and postponement of judgment. Incidence of these does not differ across groups of subjects, although we cannot infer whether more interpretations are dealt with in one way than another. It is shown only that all groups of subjects have these strategies for selection among competing interpretations of clinical data at their disposal to equal degrees. Implementation of one or a number of these strategies, it is suggested, may be dependent upon a number of factors operating at the time of decision. These may derive from the problem solver, his interpretative response, or the problem situation itself. It is agreed that, ultimately, selection, rejection or postponement of judgment must be on the basis of congruence between observed and expected information, but definition of the specificity of this process is deferred. It may be suggested, however, that the decision to accept, eliminate or postpone is the ultimate judgmental component of the diagnostic thinking process, involving evaluation of data, acceptance of the established approximation to the expected findings, the final "intellectual resolution of possibilities and actualities" discussed by Peel (1971), and the decision based, in the end, on the clinician's own estimation, according to his own criteria. These will differ since, as Peel indicates, "the decision turns on what the judger wants to fulfil". Pervading this entire process must be the inferential thinking

of the problem solver when establishing whether the observed information is actually an example of that which may be expected from the interpretations and extrapolated contexts identified, and whether or not information which has not actually been observed can be expected to be present nonetheless. Thus the three strategies of confirmation, elimination and postponement cloak a wide range of cognitive activity, but express its final dénouement.

11.7 Hypothesis 15

"Given that all subjects commence their clinical enquiry with a question intended to elicit the patient's presenting complaint, the course of the clinical interview can be determined by the following factors:

- a) The flow of information as presented by the patient.
- b) The flow of information as elicited by the subject according to his interpretations of the clinical information.
- c) The logical structure of the standard (taught) clinical history.

No differences will be observed between students, house officers and registrars in their use of (a), (b) and (c) in determining the course of the clinical interview."

Hypothesis 15 is intended to provide information and discussion relevant to the structure of the clinical interview. The Michigan four stage general model of medical enquiry is entirely hypothesis orientated, not accounting for the routine or standard aspects of the process of clinical enquiry, and so does not provide any insight into the broader issues related to the traditional structure of the clinical history. The McMaster studies do tackle this aspect of the enquiry process, finding that 61 per cent of physicians' questions are not routine while 50 per cent of physicians' and students' questions are specifically for hypothesis testing (Barrows et al, 1978; see 3.1.2 above). The McMaster group find that routine questions are used for scanning, building rapport and to gain thinking time. The differentiation between routine and non-routine may

be spurious in many instances, since questions may be identified as routine or not in relation to the intention of the poser. But it is, nonetheless, a useful frame of reference.

The categories derived from the process of content analysis for Hypothesis 15 are defined in section 10.1.6 above, as follows: category (a) is "Patient-determined interview structure", defined as "where the course of the interview as directed by S is determined by or follows on from the flow of information as presented by the patient"; category (b) is "Subject-determined interview structure", defined as "where the course of the interview is determined by the subject's requirement actively to test his interpretations of the clinical information"; and category (c) is "Logically-determined interview structure", defined as "where the subject conducts, or attempts to conduct, the interview according to a routine format as defined by the standard (taught) clinical history or any of its component parts. Exclude the presenting complaint since all subjects begin with this area". The standard history is defined in 10.1.6 above. It must be noted that, despite the category titles, by definition of role and practice, the structure of the clinical history is always determined by the interviewer, not the patient, although the patient, clearly, plays a role in determining content of the interview. Given that the interviewer determines interview structure, the three categories of Hypothesis 15 define factors which influence his decisions about that structure. In other words, the interviewer will decide (however actively or passively) on the course of the interview on the basis of the elements identified in categories 15(a), (b) and (c).

Results of statistical analysis of categories show that Hypothesis 15 is substantiated and no statistically significant difference is found between the three groups of subjects in their use of the responses indicated in categories (a), (b) and (c) to determine the course of the clinical interview. This is not to say, of course, that each student interview

might not be predominantly determined by category (a) or category (c) responses, while registrar interviews might be predominantly determined by category (b) responses. The data do not indicate results of this order, but they do show that all three categories of response are in the student repertoire of behaviour as much as the house officer or registrar repertoires. Relative use of categories by each individual may change with experience, but the potential behaviour is evident during the final year of undergraduate medical school.

The second finding of Hypothesis 15 is that, although groups do not differ, their relative use of the three categories of response does do so to a statistically significant degree (see 10.2.6 above). Inspection of frequencies per category (Table 10.13) shows that category (a) (Patient-determined interview structure) is very infrequently used. Only six per cent of all responses (or 12 per cent of all subjects) allowed the course of the interview to follow on from the flow of information as presented by the patient. Examples of such a response, where it did occur, are: "If a system came up, I dealt with it there instead of waiting for the systematic enquiry", "I decided to do the CVS there because it was relevant to what she just mentioned" and "I went on to the gut because that's what he seemed interested in". In contrast, 62 subjects made category (b) responses and 60 subjects made category (c) responses. Ninety four per cent of all responses were in these two categories. We conclude that all subjects conduct the interview according to the demands of either their own interpretations of the information or the requirements of some standard form of interview structure and content or, as is usually the case, according to some combination of both criteria. These results indicate quite clearly the problem solver's need to control the flow of information according to his own designs and, by implication, the strength of the active processes of interpretation and structuring.

Elstein et al (1978) find that diagnostic accuracy is related to thoroughness of cue acquisition and accuracy of cue inter-

pretation. On logical grounds alone, this would seem likely, but the present results add to this finding the important feature that manner of acquisition and interpretation is a prominent aspect of the process of clinical enquiry determined by the structure and content of the diagnostic thinking process. We may infer that thoroughness of data acquisition alone, in the absence of active direction by the subject's interpretative needs, may not be an efficient and effective strategy. The active role of the problem solver seems to be the most important feature of the findings relevant to Hypothesis 15.

However, having established this active, determining role, it should be considered further, since, after all, it is the patient who presents the information and, ultimately, determines its precise content even if its frame of reference is determined by the problem solver, whether clinician or student. Thus there would appear to be a reciprocal relationship between the subject's interpretative, structuring and controlling activities and the structure of the information as it is presented in a certain order, with certain degrees of clarity and specificity. This information, in turn, will effect some response in the subject and will alter his interpretations accordingly. In summary, there is a reciprocal relationship between, on the one hand, the subject's cognitive activity of structuring and interpreting the field of information, of attempting to uncover expected or suspected features or of trying to locate interpretable features, and, on the other hand, the changing field of information which, in turn, is structuring the subject's thinking processes and content by presenting new information for interpretation. Thus, the problem solver may ask a question designed to elicit a certain response, but he cannot control the answer given. That answer may confirm his interpretation of the situation, but it may also cause re-interpretation and restructuring as we have already shown in section 11.5 above. The cognitive processes of the clinical problem solver are therefore simultaneously responsive and determinative. This is so, regardless of whether the subject

chooses to conduct the interview according to the needs of his interpretative and structuring responses or the elements of a routine enquiry.

To conclude the discussion of Hypothesis 15, we may return to Barrows et al's findings concerning subjects' perceptions of the routine history as a device for scanning, building rapport and gaining thinking time. No content analysis was presently performed in relation to this point since subjects were not encouraged to voice their thoughts about it. It was therefore a matter of chance if it was discussed in more specific terms than merely identifying that routine enquiry was being made which was, in principle, the only requirement for a category (c) response to be recorded. In the event, 13 subjects commented more fully on their use of the routine history and their comments reflect the findings of Barrows et al (1978). Figure 11.5 summarises their comments. These do not purport to be representative or generalisable, but merely relevant and interesting.

Figure 11.5

Subject

Comment

S. 02

"I was just going through the standard list of diseases which ... must be asked about ... It's a bit robot-like, but if you don't ask these sort of questions, people tend to forget about them."

"He was getting my systems all mucked up. I tried to talk to him about his lungs and things and he came up with something to do with his gullet. He was getting my systems mucked up. That had to wait."

S. 06

"I'd got the idea of what the pain was, so I'd start on the old list. In the routine ... I see if I can pick up anything I might have missed."

Figure 11.5 continued

<u>Subject</u>	<u>Comment</u>
S. 17	"Weight loss and appetite loss are basic questions I would ask anybody ... that's an automatic ... The other thing about general questions is that they can be helpful in that they give you time to think, and they provide you with a reason for asking more questions that you would have forgotten about."
S. 20	"... routine enquiry in case there was anything else in the history he might have had."
H.O. 01	"I didn't actually go through all those. It depends on how lost you are."
H.O. 02	"Screening questions we usually run through ... That's tradition."
H.O. 12	"That's just a general screening. It's what the houseman's job is to do - pick up the incidental things. If you were in a clinic and you were a consultant cardiologist or something, you wouldn't ask half those things."
H.O. 16	"I'm just doing a general search. It's part of the general screening for the registrar and S.H.O."
H.O. 17	"I was stuck. When I run out of questions I go on to routine ones. It gives me time to think."
R. 07	"I was just filling in, really, the medical clerking details which we routinely do in the notes."
R. 09	"These are routine questions. Not one to be caught napping, I just wanted to make sure there wasn't any associated problem."
R. 15	"Just general background knowledge. Collecting my thoughts."

Figure 11.5 continued

<u>Subject</u>	<u>Comment</u>
R. 16	"Just systematic questions to make sure I didn't miss anything. It's just a sort of net to put everything in".

Figure 11.5 Subjects' Comments on their Use of the Routine
Clinical History

11.8 Hypothesis 16

"On review of the clinical interview, students, house officers and registrars will identify areas of omission in the information elicited from the patient. Such omission may be in two areas:

- a) Specific enquiry directed at the patient's problem, symptoms and signs or arising from the subject's interpretation of the clinical information elicited.
- b) General or routine enquiry."

Elstein et al (1978) report that diagnostic accuracy is related to thoroughness of cue acquisition. At the same time, they report that errors arise in the clinical problem solving process due to excessive data collection. Barrows et al (1978) report that their sample of final year medical students elicited 69 per cent of available clinical information with a diagnostic accuracy rate of 67 per cent; while their sample of physicians elicited 64 per cent of the available information with a diagnostic accuracy rate of 100 per cent. These findings would allow the reasonable inference that a certain threshold of information is required for diagnostic accuracy, but is not a sufficient condition. Clearly, the interpretation of that information, as Elstein et al (1978) also point out, is an equally important factor. The purpose of Hypothesis 16 is not to analyse further the relationship between information elicited and diagnostic accuracy (which came within the domain of the questionnaire study) but rather to consider the efficiency of the active process of data collection during the clinical

interview. An indication of efficiency in practice can only be gauged by reference to the subject's own theoretical potential efficiency. We are not here comparing the absolute theoretical efficiency of the perfect clinical problem solver with the actual efficiency of the subject. Instead, we are comparing the subject with himself, in practice and in theory, by noting his own judgments of his data collection during the clinical interview. If, in his account, a subject indicates that he failed to make a relevant enquiry of the patient, we may attribute that failure to variables associated with having to "think on his feet" during the clinical interview. The failure may not be attributed to his general level of knowledge, since this will not have changed between interview and account giving. We may therefore compare students, house officers and registrars on equal terms.

The categories of Hypothesis 16 are identified as: (a) Failure to make specific enquiry; and (b) Failure to make general enquiry. These categories reflect parts (a) and (b) of the hypothesis. The definitions and indicators are given in 10.1.7 above. Results given in 10.2.7 above show that incidence of responses by any group in either category is not great, ranging from four to 54 per cent. The greatest frequency is seen in the students' failure to make specific enquiry. However, the groups do not differ statistically significantly in frequency of response per category. All groups retrospectively identify failure to make a specific enquiry more often than they identify failure to make a general or routine enquiry. However, frequencies in all categories are small for both groups. Although it is not statistically significant, the 54 per cent frequency of students in category (a) (failure to make specific enquiry) requires consideration. This does indicate a tendency to have difficulty in thinking of the correct questions to ask during the interview in relation to the patient's symptoms and signs or the subject's own interpretation of these features. It may be justifiable to infer a difficulty in retrieving the appropriate knowledge, or, as Elstein et al (1978) describe it, "to select smoothly from a battery of questions and manoeuvres stored in memory". This, in turn, may have implications for

the way in which students have learned and stored information and for its relationship to the structure of information elicited during the clinical interview and the structure of interpretations made of that information. If students fail to make relevant specific enquiries during the clinical interview which, upon review in less pressing circumstances, they can identify this may well suggest retrieval difficulties related to storage structure. With clinical experience these structures may become more practically relevant and retrieval difficulties may therefore wane.

11.9 Hypothesis 18

"Results will not indicate mechanisms of actual or potential error in the diagnostic thinking process."

Elstein et al (1978) identify possible errors in the clinical problem solving process as premature closure, inappropriate selective information gathering, biased interpretation of data, mistakes in combining evidence, mis-interpretation of single cues, faulty hypothesis generation, over-interpretation, under-interpretation and uninterpretation of cues, and excessive data collection. Barrows et al (1978) cite specifically the tendency to favour some hypotheses as a source of potential error. However, despite the current hypothesis and subsequent discussion, the usefulness of identifying such sources seems limited and prone to cause attribution of undue prominence to rare phenomena. More importantly, however, there appears to be an associated logical difficulty which is that any aspect, process or strategy of clinical problem solving is inherently a potential source of error. Similar processes may result in accurate or inaccurate conclusions, depending upon the knowledge and experience of the problem solver as well as the efficiency of implementation of the process itself. However, some processes may be identified which appear to be more error prone than others. One such has already been discussed as Hypothesis 16 which concerned failure to make certain enquiries. The following discussion identifies and gives examples of other errors identified in the thinking of the students, house officers and registrars of this study. The findings of the Michigan and McMaster groups are substantiated and augmented. However, it must be stressed that the errors

to be discussed are not systematic and often are rare. Some indication will be given of the number of subjects who are identified as making the error in question. It must also be stressed that, although the errors discussed were identified in the sample of subjects, in many cases they were also rectified during the course of the interview. Finally, the discussion considers potential as well as actual error and, again, some instances (for example, designation of irrelevance) may be sources of accuracy and efficiency of thinking as well as inaccuracy, depending upon the knowledge, skill and cognitive processes of the problem solver.

Rather than attempt to classify the errors made by the present subjects according to the definitions of the Michigan and McMaster groups, we shall discuss them in their own right and relate them to the Michigan and McMaster findings.

11.9.1 Set

An error inherent in the process of interpreting information, relating it to an extrapolated context, structuring it in a certain way and conducting further enquiry according to this structure is that all information will be seen in the light of the interpretation made and flexibility to re-interpret, restructure and find new contexts will be lost. This phenomenon or aspects of it, have been identified by the Michigan and McMaster groups as inappropriate selective information gathering, biased interpretation of data, other interpretative errors and the tendency to have favoured hypotheses. Ten of the present 66 subjects made this error; these were six students, two house officers and two registrars. Figures 11.6, 11.7 and 11.8 give examples from each group. The phenomenon is well documented in the psychological literature and is usually referred to as 'set' which may be defined as "an organising, controlling and directive influence" or, more specifically to this case, "the tendency to go on performing an activity beyond its proper task-situation - a 'fixity' or over-persistence in a general direction", (Thompson, 1959). The psychological context is more fully discussed in Chapter Twelve (see section

12.2.2). However, the effect in the present context is characterised by a tendency to rationalise, ignore or find explanation for incongruent information so that the array elicited fits the identified extrapolated context.

<u>Information Elicited</u>	<u>Subject's Response</u>
Evidence of myocardial infarct	Myocardial infarct
Dyspepsia Gastric reflux	Not cardiovascular "Might be something else". Not pursued
Indigestion	Indigestion/early signs of ischaemia
Vomiting	Not pursued
Stiffness in left arm	"Not relevant to the patient's current problem "
Sickness, vomiting	"Niggling feeling that the sickness and indigestion might be due to his heart pain rather than straight indigestion "
<u>Figure 11.6 Example of Set for Cardiovascular Disease after Diagnosing Myocardial Infarct with Consequent Failure to Pursue Symptoms in Other Systems or Identify Them as not Cardiovascular. Actual Diagnoses: Myocardial Infarct, Duodenal Ulcer, Cervical Spondylosis. Subject: S. 06</u>	

<u>Figure 11.7</u>	
<u>Information Elicited</u>	<u>Subject's Response</u>
Bladder trouble Frequency, pain Blood in urine	Recent? Long term?
Stone removed from mouth of bladder, three years ago	Stone and tumour
Incontinence Sharp, burning pain Jelly in urine	Chronic problem. Tumour and stones.

Figure 11.7 continued

Information Elicited

Subject's Response

Tests show pernicious
anaemia

Anaemia secondary to malignancy
"This is just about the blood. I
was thinking what to ask her
next when she started telling
me about the tests which didn't
really fit in anywhere. But it
was helping her feel that she
was talking about what she
wanted to talk about. What you
are interested in is very often
not what the patient is inter-
ested in "

Not usually tired
Weight loss
Poor appetite

Weight loss suggests tumour,
although poor appetite could
explain it

15 minutes 29 seconds of routine history,
giving no relevant information.

Subject's Conclusions

"This is a 78 year old lady with a long standing history of
frequency of micturition, pain and, more recently, pain on
micturition and blood in her water. She's passing very
small amounts of urine very frequently during the day and
night. The pain is mainly terminal. She says she's anae-
mic and seems to have some constitutional disturbance in
that she does seem to have been tired and weak, more
recently so than in the past. The differential diagnosis,
I think it's most likely to be a tumour (carcinoma)."

Figure 11.7 Example of Set for Bladder Tumour (Carcinoma)

Causing Failure to Accept Significant Information (Pernicious
Anaemia) and Make Relevant Inquiry (Heart Failure). Actual
Diagnoses: Recurrent Urinary Tract Infection with Calculi,
Mild Heart Failure Secondary to Pernicious Anaemia.

Subject: H.O. 21

<u>Information Elicited</u>	<u>Subjects' Response</u>
20 months diarrhoea First intermittent, now continuous Tense looking patient 37 years old	1. Functional diarrhoea because patient looks tense 2. Inflammatory bowel disease Coeliac disease Anything from carcinoma on.
<u>Originally</u> , 3 motions daily No pattern Worse in morning	Functional diarrhoea occurs in the morning N.B. S ignores motions at night
<u>Now</u> , 3-4 motions at night, 3-4 motions during day	
Weight loss Normal appetite Brown, offensive motions Flush away easily No slime Blood in motions occasionally	Against steatorrhoea, inflammatory bowel disease, ulcerative colitis. Moving away from small bowel towards colon. N.B. Blood in motions and weight loss ignored
Stomach pains, but not with diarrhoea No foods affect pain Not previously unwell Blood in motions	Functional diarrhoea because pain can accompany it, but blood should not Blood could be haemorrhoids. N.B. Explains blood as not related to the main complaint. Ignores abdominal pains
Gets up at night	Not usual with functional diarrhoea but "it could mean he started off with functional diarrhoea and has developed something else " N.B. Information not accepted as excluding functional diarrhoea
No migraine No tablets No worry with job Worry outside work, not associated with diarrhoea Bran and neomycin helped	"I've learnt that somebody else thought it was functional and has given him bran " Worries outside work is "a little more evidence on one side of the balance " N.B. Response to antibiotics ignored, effect attributed to bran alone
<u>Figure 11.8 Example of Set for Incorrect Diagnosis (Functional Diarrhoea) Causing Misinterpretation and Ignoring of Significant Information. Actual Diagnosis: Crohn's Disease</u>	
<u>Subject: R. 14</u>	

Figures 11.6, 11.7 and 11.8 illustrate a range of cognitive processes which enable the subject to maintain his current interpretation in the face of apparently opposing information. These examples do not indicate the psychological reason for such strength of interpretation or resistance to flexibility of interpretation. However, it may be said that in these cases the identification of an extrapolated context and consequent expectancy of specific concomitant information precludes the subject from evaluating each item of new information in its own right. Instead, all information is evaluated solely for its relationship to expected findings. If it does not fit in, then it is either ignored or an elaborate explanation is constructed for the finding. Evaluation of information in its own right may, therefore, be an important cognitive skill to be taught at both undergraduate and postgraduate levels of education and training.

Some subjects do show awareness of this potential error, and make special efforts to avoid it. For example:

"I was trying to making myself think of other possibilities, because I do tend to get a rather fixed idea about one thing, and exclude everything else, and to pursue that particular idea without thinking of a differential diagnosis."
(S. 17)

"You can't ask them unbiased objective questions. You find you're looking for specific answers ... so you put more emphasis on some questions and less on others because you've already presupposed the answer." (H.O. 20)

However, successful avoidance of the error of set requires a number of cognitive operations: evaluation of information both within and independently of the extrapolated context already identified; stored information structures which permit easy and appropriate access at a variety of points; cognitive flexibility in interpretation and re-interpretation of information; self-awareness.

11.9.2 Failure to Make the Correct Interpretation

Failure to make the correct interpretation of information may

be for one of two reasons: either the problem solver does not have the pre-requisite knowledge and therefore cannot make the correct interpretation of information; or, he fails to find the correct extrapolated context for some reason other than lack of knowledge. It is this latter case which concerns us here since we are interested in errors of thinking, not inadequacy of knowledge per se. Figures 11.9 and 11.10 give examples of unexplained failure to make the appropriate interpretations of information elicited. In each case the actual diagnosis (asthma and tuberculosis) is not obscure and we may be certain that the subjects possessed sufficient knowledge to make the correct diagnosis. In both cases, also, it would appear that sufficient information was actually elicited to do so.

In explanation of the subjects' error (which occurred in six of the sample of 66 subjects), we may again refer to the

Figure 11.9

<u>Information Elicited</u>	<u>Subject's Response</u>
Cattarrh Can't breath properly Intractable cough Doctor said bronchitis	Need to clarify cough
Not productive Blood once or twice Breathlessness Cough for 6 months	Respiratory/cardiovascular/ carcinoma in lung/primary cardiac failure/left vent- ricular failure
Three pillows Stopped smoking Tailor	"
Several visits to GP Two hospital admissions	Probably not carcinoma
No operations No other illnesses GP said rheumatism (No question relevant to asthma)	"
No TB No diabetes No relevant family history	"

Figure 11.9 continued

<u>Information Elicited</u>	<u>Subject's Response</u>
Left-sided chest pain Coughing Breathing difficulty	Cardiogenic cause/myocardial infarct/angina
Many treatments Allergic (N.B. does not pursue this) GI tract normal	Cardiac/pulmonary cause of cough which is the main complaint

Subject's Conclusions:

"He's a man with a cough which is from time to time associated with left-side chest pain, and he's short of breath, sometimes in the middle of the night. There are many causes of shortness of breath, largely, they're either pulmonary or cardiovascular. Having been through the whole lot of symptoms with him, I would have plumped for cardiovascular cause."

Figure 11.9 Example of Failure to Extrapolate to the Correct Context. (Asthma). Actual Diagnoses: Intrinsic (late onset) Asthma, Acute Bronchitis and Allergic Aspergellosis.

Subject: R. 09

related concepts of interpretation and information structure. It can be seen from the two examples given that the set effect is not operating to cause a persistent misinterpretation, yet the subjects' response to information still does not enable them to make the correct extrapolation. We may only speculate about reasons for this, but may state quite certainly that, since all necessary information would appear to be available, it is the process of structuring that information appropriately which fails. But what could 'structuring' mean in this context?

Structuring in the diagnostic thinking process almost invariably requires the problem solver to select significant items of information from among an array of relevant and irrelevant items and to inter-relate them appropriately according to some criterion or criteria. The hazardous effects of large quantities of

<u>Information Elicited</u>	<u>Subject's Response</u>
Swollen feet	Heart failure/nephrotic syndrome.
Shortness of breath	Heart failure
Fits	
SOB for 6 to 7 months	Heart failure/chronic bronchitis/bronchitic disease giving heart failure
Aged 56	
Fit before SOB	
Decorator	
No chest pain	All against heart disease
No stomach pain	
Lost appetite	
Dry cough	
Two pillows	
No SOB at night	
No abdominal swelling	No ascites/no liver disease/
No bronchitis	no cirrhosis/renal disease/
Used to smoke 10-12 daily	no lung disease
Guinness most days	
Ankles better	Varicose ulcer, badly infected
SOB gone	"I really can't think of any
Previous month in isolation ward	associates that would give him
No water tablets	ankle oedema and being on an
	isolation ward "
No varicose ulcers	Not varicose ulcers
No other illnesses	Down and out
Married	But married
Angular stomatitis	
<u>Figure 11.10 Example of Failure to Extrapolate to Correct</u>	
<u>Context. Actual Diagnosis: Pulmonary Tuberculosis with</u>	
<u>Hypoalbuminaemic Oedema. Subject: R.21</u>	

irrelevant information have been discussed in 3.3 above. However Figures 11.9 and 11.10 do not seem to suggest that information overload or irrelevance is the problem. Rather, it would appear to be a failure to identify the true forceful feature and structure the related information accordingly. In rectification of this error, then, we are brought back to the set of necessary cognitive operations identified at the conclusion of

11.9.1 above concerning evaluation of information, cognitive flexibility and self-awareness. Information being elicited can only be interpreted and structured according to the limitations of the subject's own information store in both its structure and its content, which, in turn, have direct implications for its accessibility (or transferability). It is the flexibility of access to the stored structure which will eventually determine the efficiency and effectiveness of the diagnostic thinking process, given an adequate knowledge base. Such flexibility may be partially related to a preliminary structure which enables identification of forceful features from among an array and extrapolation from there to possible embedding contexts.

11.9.3 Designation of Irrelevance

It has already been pointed out (see 11.9 above) that designation of items of information as irrelevant may or may not be appropriate, since in most clinical histories it is likely that some information actually is irrelevant to the diagnosis. However, in some cases, information may be designated irrelevant when it is, in fact, of diagnostic significance. Inappropriate designation of irrelevance may be for one of two reasons; either the subject has inadequate (incorrect or insufficient) knowledge, or his problem solving thinking process is at fault. It is the latter case which is of primary interest for the current discussion. Examples of this may be taken from the subjects of the study:

"I didn't ask her why she fell over ... because it didn't seem that relevant to what she's in and being treated for." (S. 04)

"I couldn't make anything of that symptom, so I completely pushed it out of my consciousness. Intuitively, I suppose, I assumed that it wasn't anything particularly relevant." (H.O. 02)

"I find it rather difficult to evaluate all his different pains. There's nothing much you can do except accept that he's got them ... I shall leave them there." (H.O. 03)

"The shoulders were a bit puzzling. I didn't expect that. Referral through to the back was fine. I didn't know what to make of it, so I dismissed it." (H.O. 08)

"... but the deafness; I'm not a neurologist; I can't remember what's nerve deafness and what's bone deafness. I just discarded it straight away." (R. 02)

"There are lots of bits that just don't fit in, so I just had to skate over them." (R. 05)

"Just going along with the red herring of thyrotoxicosis to see if she's got any diarrhoea or heat intolerance, which she has. But it's over the last 18 years, so it's pretty irrelevant really." (R. 04)

These examples show a variety of reasons for designation of irrelevance which each may be classified as either failure to incorporate into an extrapolated, structured context, or failure of immediate knowledge. In some cases these two may be inter-related or may even appear to be the same thing. In summary, however, it may be said that, as with previous errors, structuring and interpretative aspects of the diagnostic thinking process are heavily implicated where the knowledge base is adequate. Where the knowledge base is inadequate, and recognised as such, the routine history provides a framework for thorough data collection, at least. However, where the problem arises from inability to locate the appropriate context in which the information elicited may be found embedded, then solutions can only be found in terms of the same factors as those identified in the conclusion to 11.9.1 above.

The error here identified as designation of irrelevance is comparable with Elstein et al's error of uninterpretation. Related to this are errors of over-interpretation and misinterpretation. It would seem reasonable to suggest that these complementary errors may all be attributed to the same structuring mechanisms of thinking as those here associated with the designation of irrelevance. This latter is a potentially useful skill, unlike the complementary ones, but it is

a skill which should be employed consciously and in a reasoned manner. The alternative mechanism must be, as the McMaster students are advised (Learning Resources Design Project, 1975), to generate a 'net' of hypotheses that will encompass all items of information. Such a process would result in the generation of irrelevant hypotheses, thereby overloading and decreasing the efficiency of the hypothetico-deductive system itself. Designation of irrelevance is a useful skill and one which potentially enhances efficiency. It must be applied, however, with conscious caution.

11.9.4 Conclusions

The review of actual and potential error in the diagnostic thinking process (which must include the discussion of Hypothesis 16) admits of one main, generalisable conclusion of practical value to either medical education or clinical practice. The conclusion must be that potential errors in the process, apart from unpredictable, idiosyncratic ones, arise from either inadequate knowledge or error of the cognitive process itself. If lack of knowledge is the cause, then the problem solver can recognise this and compensate accordingly. If he is misinformed and unaware of being so, then no self-monitoring system of error avoidance is possible. However, it would seem that most potential error arises from features of the diagnostic thinking process itself. In particular, error arises either from failure to gain appropriate access to the stored structures of information and thereby failure to make correct (or any) interpretation of the elicited information, or, error arises from the comparison of elicited and expected information where flexibility of interpretation and facility of restructuring are absent. In both cases, error may be more frequently avoided by attention to strategies of evaluation of items of information both within and independently of the extrapolated contexts already identified; stored information structures which permit easy and appropriate access at varied points; cognitive flexibility in interpretation and re-interpretation of arrays of information; and, self-awareness in the dynamic diagnostic thinking process.

11.10 Hypothesis 19

"Results will not indicate any categories or types of information other than those of the standard (taught) clinical history, used to assist either in pre-diagnostic or diagnostic interpretation of clinical information or in selection of the most likely diagnosis."

Hypothesis 19 enquires whether or not subjects use information other than the direct clinical information indicated by the headings of the standard history (see 10.1.6 above). In particular, it enquires whether such information is gathered for diagnostic, rather than informative, purposes. If it is, then attention should be paid to the skill of estimating reliability of information.

Evidence from the accounts of the 66 subjects of this study indicate that types of information other than those indicated in the standard history are routinely used for diagnostic purposes. Four categories of such information may be identified: tests and procedures; treatment; circumstantial evidence; and, other clinicians' opinions.

As an example of use of information about tests and procedures, subject S. 08 restructured his interpretation of the patient's problem to encompass TB because he discovered that the patient had been investigated for this problem. Subject S. 10 extrapolated to the context of cardiac problems, despite lack of evidence, upon discovering that the patient had been investigated by echocardiogram. Subject S. 12 made the interpretation of 'pneumothorax' upon discovering that the patient had been "breathing through a bucket of water" despite lack of other confirmatory findings:

"He denied ever having any pain in the chest at all, which is a bit unusual. But at that point I thought he might have had a spontaneous pneumothorax. I think I was fairly sure about that because I can't think of anything else that you treat like that".

As one subject explained:

"It's a technique doctors use when we're not sure of the diagnosis. If you find out what tests were done, you can find the diagnosis much easier." (H.O. 05)

Treatments were enquired about in the same way: names of antibiotics given, "definitive therapy", treatment in a coronary care unit, special diets, operations. All were used as diagnostic information: "I wanted to know what cleared it up, because that tells you what caused it" (S. 15). Likewise, circumstantial evidence of ward, special unit, the surgeon's name, the consultant's name, knowledge of the clinician's special research interest or style of patient management were all treated by students, house officers and registrars as valuable, diagnostic information. One subject explained:

"You have to get every single item of evidence out, particularly what other doctors told him, what his GP told his wife, what medicines he was given and what his response was to it." (R. 06)

The category of other clinicians' opinions was treated with caution only by a registrar who discouraged his students from believing such reports and by a house officer who felt that you really could not believe patients' reports of what other doctors had said.

In conclusion, it appears that information other than that indicated by the headings of the standard history is used by all three groups of subjects. The frequency of that use cannot be determined, although 35 of the 66 subjects of this study made some reference to use of such information. This being so, it would seem reasonable to suggest that consideration be given at some point in the teaching and learning process to factors in the reliability and validity of information, in relation to its relatively direct or mediated nature.

11.11 Hypothesis 20

"Results will provide no indication of the nature of psychological probability in the diagnostic thinking process."

Chapter Two presents a discussion of mathematical and logical models of the diagnostic process and concludes that these are paramorphic at best, and certainly not isomorphic. A number of these models are probabilistic and ideas of numerical probability as part of the diagnostic thinking process are

questioned and rejected. However, in rejecting numerical probability, the concept of psychological probability is not also eschewed. Indeed, the associated concepts of diagnosis as an ill defined problem, of judgment and inference and of partial knowledge lead to the inescapable conclusion that uncertainty is endemic in the diagnostic process. Uncertainty in medicine is the proper study of some workers (Cohen, 1972; Clouser, 1977). By definition of process, each clinical problem is unique yet to be identified as a special case of a general principle. Clinical information is often indirect, relying on reports of subjective experience. The question of how a diagnosis is finally selected, therefore, not only concerns the categories discussed in relation to Hypothesis 14 which concerns selection among competing interpretations (section 11.6 above) but also presupposes a judgmental component which may be considered in terms of psychological probability. In the event, the present data present little information of explanatory or descriptive value.

Subjects tend to weigh up and balance data, assigning relative importance to contradictory items and reaching a conclusion with some degree of definiteness. We are reminded of Elstein et al's (1978) 'judgmental rules' (see 3.1.1 above) which balance positive and negative items and reach a conclusion accordingly. For an example of the process, we may cite a student:

"The picture's slightly atypical for angina. The pain described is a shooting pain not a constrictive pain. But it does radiate down the arm, and she's known to have heart disease. So I think it is angina. And it's also related to exercise. So it is angina. I think so." (S. 16)

A registrar gave an account of the same type of thinking process:

"... the one point in his history that makes me think it's ischaemia of the small bowel over every other diagnosis, is that he has his symptoms relieved very promptly by posture. This is something which is, perhaps, not in keeping with hiatus hernia or a duodenal ulcer. And, remember, he had a duodenal ulcer in the past and he says

his present symptoms are like that. But his other symptoms ... It's not precisely clear, I don't think, as to what it is. One might expect to see some abnormality of the bowels ... but not necessarily so. So, primarily I think it's ischaemia of the gut. Why? How could we relate this to his other symptoms of palpitations, chest pain, dyspnoea? Well, he could have had an embolus from his left ventricle or left atrium which would have settled in his gut and caused ischaemia. Out of the two-paroxysmal atrial fibrillation and paroxysmal sinus tachycardia, I would favour the latter because of his history of polyuria after his attacks. Although polyuria is seen after paroxysmal atrial fibrillation" (R. O3)

Thus the information is weighted, balanced, fitted with an expected array, explained and considered and a judgment made. The two accounts quoted are not untypical. But they do not permit inference to the underlying process. It is clear that subjects could be asked to rate numerically or positively or negatively the items of information so considered and some formula such as Elstein et al's may accrue. Yet this does not demonstrate the isomorphism of the model, only the subject's ability to perform the rating task. For the moment, the process must remain only partially illuminated.

11.12 Conclusions: Interpretative Value of the Present Study for Current Descriptions of the Diagnostic Thinking Process

Current descriptions of the diagnostic thinking process are reviewed in Chapter Three above. The major model is that of hypothesis generation and testing. This model is elaborated in a number of ways, which are summarised in section 3.5. The only other model of any currency is that which interprets the diagnostic thinking process in terms of pattern recognition. This model is rejected as inappropriate. The current discussion, therefore, will relate only to the 'hypothesis generation and testing' model of the diagnostic thinking process and the elaborations of that model put forward by the research studies at Michigan and McMaster. The models are criticised in particular for leaving unanswered a number of fundamental questions. These are also summarised in section 3.5. The discussion of

Hypotheses 10 to 20 has addressed these questions and although the relevant findings and discussion do not purport to answer them fully, it is considered that useful frameworks have been built. Rather than reiterate the discussion already presented we may simply indicate the questions and particular sources in this chapter of relevant discussion, thus:

How are cues cognitively manipulated? See 11.5

What process of structuring generates hypotheses?
See 11.2 and 11.5

How can multiple hypotheses be generated? See 11.2

What cognitive mechanism enables multiple interpretations of cues? See 11.2 and 11.4

How do new hypotheses arise? See 11.5

What cognitive processes occur before generation of the first hypothesis? See 11.3

How are relationships between cues allocated and reallocated? See 11.5

According to what tactic or strategy is the clinical interview conducted prior to generation of the first hypothesis? See 11.3

How is information designated as irrelevant? See 11.9.3

We may now consider the elaborated model of hypothesis generation and testing in the light of results and discussion of the current account gathering study. The Michigan and McMaster models are only slightly different, but such differences as there are will be made clear.

Elstein et al (1978) postulate a four-stage general model of medical inquiry which consists of cue acquisition, hypothesis generation, cue interpretation and hypothesis evaluation. The present results and discussion question the differentiation between the second two stages (see 11.2 above). It is suggested that these are two aspects of the same process and that the process which is referred to as hypothesis generation is,

in fact, a process of cue interpretation. This conclusion derives from a consideration of the cognitive process which allows the problem solver to attribute meaning to the information elicited. Such a process must involve acting upon the information, structuring or organising it, in the process of which it is identified as potentially being an instance of X, Y or Z where X, Y or Z might represent undefined pathology of an organ ("something wrong with the liver"), undefined pathology at a certain level of specificity ("something metabolic"), a functional abnormality ("left ventricular failure"), a symptom ("paroxysmal nocturnal dyspnoea"), a disease ("chronic lymphatic leukaemia") or any other classification, disease, process, procedure and so on. Any of these will be based on the same cognitive process of, somehow, relating the information given to information stored. Whether it is a complex process of selecting, weighing up and relating symptoms and signs and identifying a differential diagnosis of some specificity, or whether it is translating "I have to have four pillows, doctor" into "paroxysmal nocturnal dyspnoea", the cognitive process must be the same, varying only in complexity. It therefore seems untenable, in psychological terms, to differentiate between hypothesis generation and cue interpretation.

In considering the cognitive process of interpreting, structuring or organising information and relating it to stored information thereby extrapolating to specific expected information or, at least, being enabled to identify questions the answers to which should allow such extrapolation, the real difficulty lies in identifying the beginning of the process. Barrows et al (1978) resort to theories of pattern recognition but this merely begs the question, since for the information array to be recognised as a 'pattern' it would have to be organised as a pattern. According to what criteria would the problem solver embark upon that process? It can surely only be done with reference to information not present, which returns us to the question of how that information is first located. As far as such location goes, the present results (see section 11.3) indicate that either pre-diagnostic or diagnostic interpretations, or both, may be made as a primary

response to clinical information elicited. This clarifies the somewhat confused McMaster findings concerning relative initial use of broad and specific hypotheses (see 3.1.2 above).

The Michigan findings are no more helpful than the McMaster ones in identifying the origin of the interpretative process. They define this as attending to initially available cues, identifying cue clusters and associating to long term memory. But such a process is as logically inconsistent as 'pattern recognition' and for the same reasons. A 'cue cluster' is only such in the mind of the clinician, when he has responded to the stimuli presented by the patient and formed 'cue clusters' for himself. Unless the problem solver acts upon and cognitively organises the information given by the patient he would in many cases, be unable to relate it to anything in his long term memory other than dealing with it item by item. This argument may lead us to consider that information is not stored in the problem solver's memory simply as a series of arrays each of which represents a pre-diagnostic interpretation, but rather that information is also stored such that the problem solver may recognise a forceful feature and make his connections from there. In other words, perhaps the process begins with identification of the item or items of information with the greatest payoff in terms of facilitating interpretative or structuring responses which will change the ill defined problem into a well defined one. Such identification of the valuable information may change with experience, as the process is reinforced or not.

Having discussed how the diagnostic thinking process begins, it is necessary to discuss the related question of where it begins also. Elstein et al (1978) state that diagnostic hypotheses are generated early in the clinical interview, and may be broad or specific. The McMaster group reach the same conclusions and add that students and physicians do not differ in time of generation of the first hypothesis. The present study shows agreement that interpretations may be broad or specific at any point during the process and elaborates on reasons for this (see 11.2) in the light of cognitive processes of structuring

and extrapolation to related contexts. In addition, the present study indicates that the diagnostic thinking process is in constant, responsive, organising and determining activity. The loose identification of 'early' hypothesis generation is inadequate in describing the initiation of the process. This inadequacy would appear to arise from the search for characteristics of hypotheses rather than a broader and more flexible analysis of a thinking process more refined, complex and varied than the characterisation of hypothesis generation and testing has allowed.

The present study shows (section 11.3) that subjects can and do make active, interpretative or evaluative response to all clinical information as soon as it is elicited. Such a process does not occur 'early', but immediately. Elstein et al's (1978) contention that hypotheses are generated from clusters of a few cues may well be correct in their terms, but it is inadequate, ignoring the cognitive activity of the clinician prior to eliciting sufficient cues to form clusters. The present study has indicated that even a single piece of clinical information can be and is structured in an extrapolated context and accordingly interpreted. We therefore refute the implicit assumption of the Michigan and McMaster groups that, at some points during the diagnostic process, the problem solver's mind can act merely as a passive register. The details of this argument are presented in section 11.3 in refutation of theories of pattern recognition in the diagnostic thinking process. Instead, it is suggested that information is responded to as soon as it is elicited and that response might be something which the Michigan or McMaster groups would recognise as "hypotheses" or it might not be. For example, it might be a judgment of need for further enquiry before satisfactory interpretation or structuring of the information is made possible. This illuminates an aspect of the interpretative cognitive processes which we suggest, concerning the problem solver's preferred methods of working or subjective judgment of methodological efficiency or effectiveness or familiarity with the extrapolated contexts. Thus it is likely that the problem solver could extrapolate from any given piece of information,

but the evaluative category of judgment of need for further clarifying enquiry, suggests that he is more or less willing or able to work within the resultant structure than another possible, alternative one.

Theories of the diagnostic thinking process as hypothesis generation and testing do not account for such cognitive phenomena as this. We therefore prefer a broader, psychological definition and explanation of the diagnostic thinking process in terms of structure and extrapolation on the basis of assigned and stored information structures. Such a definition subsumes within a broader and more precise context the activities identified as hypothesis generation and testing. This is discussed in section 11.4.

We have, therefore, identified and described the manner and point of origin or initiation of the diagnostic thinking process. The dynamic of the process, once underway, is not specifically addressed by the Michigan and McMaster groups, but is discussed here in section 11.5 in terms of the mechanism of interpretation and reinterpretation of the changing array of clinical information throughout the interview and diagnostic process. However, the North American groups do identify numerical "rules" for the termination of the process, that is, for selection of the most likely diagnostic hypothesis. These our results and discussion have led us to reject (section 11.6), not for their probable artefactual nature, but for their implicit simplistic model of the variables inherent in the diagnostic problem and process. However, we would agree that the problem solver does work by comparing observed and expected findings, but that this process, again, is not as simple as Elstein et al (1978) seem to imply. Information is often evaluated in its own right before being compared with an expected model (see the example given in 11.6) and it seems unlikely that final selection of the most likely diagnosis is made on the basis of numerical pre-dominance of items of information related to one rather than another. It is suggested (11.11) that numerical and psychological probability are very different phenomena.

The present findings occasion some reconsideration of the assumption of the hypothesis generation and testing model that

the latter activity (hypothesis testing) determines the course of the clinical interview. It must be recalled that the McMaster group do discuss the role of the routine history and section 11.7 indicates our own agreement with their findings. However, to return to the determinants of the course of the clinical interview, we find that hypothesis testing is too undifferentiated and nebulous a description. The course of the interview may be determined by the subject's response to the flow of information as presented by the patient, or the logical structure of clinical histories as epitomised by the routine history, as well as by the problem solver's own interpretations of the information. In addition, there is a subtle and complex reciprocal relationship between the problem solver's cognitive activities and the information as it is presented with order, clarity and specificity uncontrolled by the enquirer. This is discussed fully in 11.7.

Elstein et al (1978) and Barrows et al (1978) identify errors in the clinical problem solving process as described in section 11.9. Their findings are substantiated by the present results in terms of the set effect, failure to make the correct interpretation of the information, designation of irrelevance and failure to make an appropriate enquiry. However, the appropriate context of such errors is defined in the current discussion by drawing attention to the logical inference that any aspects, process or strategy of clinical problem solving is inherently a potential source of error depending upon the thoroughness and rigour of its implementation and the accuracy and completeness of the problem solver's knowledge.

Finally, two points must be mentioned. Firstly, the hypothesis generation and testing model has given rise to specific numbers of hypotheses being attributed to subjects at any one time. This is questioned on both theoretical and empirical grounds (see section 11.2) as misleading and limiting to analysis of the cognitive process which facilitates such multiple interpretations of the same information. Secondly, it is worth noting that in this entire study, no differences have been found between the range of cognitive processes of students, house officers and registrars. This is not to say that within any one

person one may dominate over another or be preferred for some reason related to experience and practice. Despite this possibility, the finding of no differences is something which must be dealt with in the light of criticisms of medical education for its lack of teaching in the area of diagnostic thinking skills and which suggest that such skills develop only with clinical practice (Barrows and Bennett, 1972; see also Chapter One).

In conclusion, the interpretative value of the present study for current descriptions of the diagnostic thinking process may be summarised. Although theories of pattern recognition are rejected on both theoretical and empirical grounds, the pre-dominant theory of hypothesis generation and testing cannot be dismissed quite as summarily. As a complete and accurate description of the clinical problem solver's thinking processes, it is clearly inadequate. It is also assessed as unnecessarily limiting and misleading. Given the small part of the diagnostic thinking process to which it is addressed, it may seem a reasonable description of that portion of the clinician's behaviour. However, if the wider process is considered and serious questions asked about the nature, derivation and fate of hypotheses, about the psychological rationale and meaning of such a process, its origin and mechanism, the inadequacy and inaccuracy of the formulation becomes apparent, as does its tendency to deflect attention to superficial features of thinking rather than underlying processes. Likewise, the variables or identifiable aspects of the diagnostic problem and thinking process have not been assigned a role in the theory. For example, we may cite variables of information flow, determinants of interview structure, characteristics of elicited and stored information and so on. We therefore suggest that the more psychologically precise explanatory concepts of structure, interpretation and extrapolation be employed instead of the bland and generic theory which hypothesis generation and testing seems to be. Such concepts have the necessary characteristic of giving rise to testable research hypotheses and questions and, as is discussed in Chapter Thirteen, associated specific principles from which to develop appropriate teaching and learning strategies.

11.13 Summary

The conclusions of this chapter may be summarised in point form as follows:

1. Two forms of interpretative response to clinical data are identified. These are pre-diagnostic and diagnostic interpretations. The former may be seen as working interpretations only, not hypothesised diagnoses.
2. Multiple interpretations of both types are made by all groups of subjects who appear to have equal interpretative capacity.
3. Multiple interpretations of the same array of information can arise only if that information is either cognitively rearranged to have a variety of possible internal structures and inter-relationships or if it is related to many different extrapolated contexts of information structures and stored in the subject's memory.
4. The number and nature of interpretations possible must be limited and determined by the information elicited as well as the structure, content and accessibility of the subject's store of knowledge.
5. The problem solver has the ability to structure and interpret information in ways other than those which he is currently considering at any one time. Thus, the diagnostic thinking process is both creative and dynamic.
6. Students, house officers and registrars do not differ in their relative use of pre-diagnostic and diagnostic interpretations.
7. Pre-diagnostic interpretations are statistically significantly more frequent than diagnostic interpretations for all groups, but the diagnostic thinking process is not one of progressive specification. The linear only partially controlled nature of the information flow from patient to clinician may pre-dispose him towards initially extrapolating to pre-diagnostic interpretations rather than diagnostic ones.

8. Subjects in all groups can and do make active, interpretative or evaluative response to clinical information as soon as it is elicited. Theories such as pattern recognition which imply passive reception of information are untenable. All information is structured and interpreted in some way and not merely gathered.
9. All groups of subjects show evidence of evaluating either their own interpretative processes themselves or their outcomes and of greater or less willingness or ability to work within one structure rather than another.
10. For all subjects, the diagnostic thinking process involves working within extrapolated cognitive contexts and defining the 'fit' of observed and expected findings. Whether the context is pre-diagnostic or diagnostic depends upon the information elicited and the structure of the subject's stored knowledge.
11. Desirability of re-interpretative and interpretative thinking cannot be judged in absolute terms, but only in relation to specific cases. However, the ability to think in such ways is desirable in the diagnostic problem solver.
12. Re-interpretation in the absence of new information is less common than re-interpretation in the presence of new information for all groups. Stimuli to each of these processes are varied.
13. Re-interpretation occurs as a result of cognitive restructuring of information and consequent embedding of the information in a new extrapolated context of possibly related information.
14. Students, house officers and registrars do not differ in their strategies for selection between competing interpretations of clinical information. Subjects tend to use a combination of strategies, with a more common use of confirmation rather than elimination or postponement of judgment. Strategy may depend upon personal factors or upon characteristics of the information.

15. Clinical data are evaluated in terms of the fit to anticipated possible findings.
16. Competing interpretations are dealt with by confirmation, elimination or postponement of judgment. Such decisions represent the ultimate judgmental component of the diagnostic thinking process.
17. Differentiation between routine and non-routine questions is hazardous.
18. All groups show a similar range of strategies for determining the course of the clinical interview, although some may pre-dominate in some groups. All subjects conduct the interview according to the demands of either their own interpretative responses or the requirements of some standard form of interview structure and content, or, more usually, both.
19. Thoroughness of data acquisition alone, in the absence of active direction by the subject's interpretative needs, may not be an efficient and effective strategy.
20. There is a reciprocal relationship between the subject's interpretative, structuring and controlling activities and the structure of the information as it is presented. The subject's cognitive processes are simultaneously responsive and determinative.
21. Barrows et al's views of the routine history are substantiated.
22. All groups retrospectively identify failure to make a specific enquiry of the patient more often than failure to make a general or routine enquiry. Fifty four per cent of students identify failure to make specific enquiry, indicating failure to think in action. This may be attributed to the structure and accessibility or organisation of stored information.

23. Any aspect, process or strategy of clinical problem solving is potentially a source of error. Special errors are identified as the set effect, failure to make the correct interpretation, and designation of irrelevance.
24. Information other than that indicated by the headings of the standard clinical history is used by all groups to assist either in pre-diagnostic or diagnostic interpretation or in selection of the most likely diagnosis. Such information may concern tests and procedures, treatments, circumstantial evidence (ward, consultant, etc.) or other doctors' reported opinions.
25. Psychological probability involves processes of weighing up, balancing and fitting observed evidence to expected possible information.
26. Psychological concepts of information structure and interpretation based on extrapolation to possible contexts are considered more helpful and accurate than the pre-dominant interpretation of the diagnostic thinking process as hypothesis generation and testing.

Having reached these conclusions on the basis of the account gathering study, we may now proceed to unite them with the conclusions of the questionnaire study in order to make a more complete definition and description of the diagnostic thinking process in students, house officers and registrars. We may also now relate these findings to general psychological theory. This forms the discussion of Chapter Twelve.

Footnotes on Page 290

- (1) Defined as the feature (or features) of a stimulus situation which, above all other features actually does evoke the response.
- (2) Defined as the actual characterising feature (or features) of a stimulus situation, as opposed to some other feature (or features), which might or might not be attributed force by the observer, but which characterises some other actual stimulus situation.

CHAPTER TWELVE

The Diagnostic Thinking Process: A Psychological Perspective Relating the Results of the Two Parallel Studies

The results and findings of the two parallel studies which comprise this enquiry have been described and discussed separately. This chapter relates and unites these separate elements in order to derive a unified and consistent explanation of the diagnostic thinking process in undergraduate medical education and clinical practice. Section 5.6.3 above establishes the comparability of the subjects of the two separate studies. We may therefore conclude that it is both reasonable and justifiable to unify the findings of the two parallel studies and so derive one theory of the diagnostic thinking process.

However, before proceeding, it is necessary to consider whether it is justifiable to compare a study which recognises specialities (the questionnaire study) with one which does not (the account gathering study). Could it not be that concealed beneath the results of the latter are gross speciality differences in the thinking processes revealed? The refutation of such an assertion must be of two types: a priori and a posteriori. The a priori argument derives from the design of the account gathering study which includes a wide range of clinical material and diagnoses and therefore yields information about general characteristics of the diagnostic thinking process. In addition, the work of the Michigan and McMaster groups had already concerned different specialities and revealed similar cognitive processes of a general type, while findings of speciality specificity have been shown to involve variables other than those of the type which could be expected to result from the account gathering study. The a posteriori argument concerns the nature of the cognitive processes demonstrated in the account gathering study data. These are processes which are referable to general cognitive development and are not particular to the diagnostic thinking process per se (see 12.3 below). It is therefore reasonable to consider that these processes are apparent regardless of the speciality concerned, although the relative contributions of those processes, as is shown, may vary. It is therefore considered justifiable to unify the results of the two parallel studies from the points of view of both the subjects involved and the content.

The theory to be presented defines the characteristics of the diagnostic thinking process and problem. The description of the stages of that process is in the broadest of terms. The unpredictability, individuality, vagaries and indirection of each clinical problem and process preclude linear or ordered description or prescription of absolutes. The following sections, then, present, firstly, a discussion of the fundamental psychological processes identified. This is followed by discussions of special features of the stages of initiation, progress and resolution of the diagnostic thinking process. A discussion of developmental and comparative aspects is then presented.

12.1 Fundamental Psychological Features: Structure and Extrapolation

The discussions of the results of the account gathering study (Chapter Eleven) and of the principles and practice of teaching the diagnostic thinking process (Chapter Four) have introduced and substantiated the centrality of the explanatory concepts of structure and extrapolation. Although closely related, each of these may be discussed separately in relation to the results of the two parallel studies and to wider psychological theory.

12.1.1 Structure

It is our contention that the diagnostic thinking process involves the active assignment of some structure to the clinical information as it is elicited. In psychological theory, the search for structure has been presented as inherent in human behaviour (Garner, 1962). Section 4.1.1 discusses the psychological and pedagogical meaning of structure, at length. In principle, however, we are referring to the ways in which pieces of information are or may be related either in reality or in the clinical problem solver's mind.

In section 8.2 are identified some sources of differences between students and registrars. These may best be understood in the light of the concept of structure. For example, the different contextual aspects of knowledge and skill, relative rehearsal of knowledge and skills and the various contextual

aspects of that rehearsal may differentiate between the diagnostic thinking processes of students and registrars primarily because of the different effects of these experiences on the structure and accessibility of stored knowledge. In turn, such differences in structure and accessibility of stored knowledge may yield differences in the structure assigned to the clinical information elicited. Results of the questionnaire study give substance to this argument. It is shown, for example, in section 8.3.1 that in endocrinology factual knowledge plays a not inconsiderable role in the students' diagnostic thinking processes, while for registrars this capacity plays a minimal relative role. This difference may be attributed to differences in the contexts of acquisition and use of knowledge. The account gathering study substantiates this conclusion by demonstrating a tendency in students to find difficulty in gaining access to or using the stored knowledge structures during the clinical interview (section 11.8). We may infer from this difficulty that these structures may not be wholly appropriate to that situation. It is reasonable to presume that students' knowledge is primarily acquired in a context other than that of the patient presenting a flow of clinical information, and that the experience of a fairly formal learning process predominates over the experience of acquiring information in the active clinical problem solving context. It is therefore to be expected that the context of a formalised knowledge structure will play an important part in the diagnostic thinking process. As that knowledge is used in more and more clinical contexts to serve the diagnostic thinking process, it is reasonable to infer that its structure alters and that the nature of the diagnostic thinking process alters accordingly. The same may be said of the role of interpretation of symptoms and signs in endocrinology as tested in section B of the questionnaire.

In both endocrinology and neurology, it is shown that the skills taught during the undergraduate course are those which play important roles in students' diagnostic thinking processes, and that these may or may not be the same as those of the experi-

enced clinician, depending upon the pedagogical approach of the speciality. In both cases it is also clear that there is development and change in the diagnostic thinking process with clinical practice. It is reasonable to infer from such similarities and differences that one aspect of change is in terms of the structure of relevant knowledge as well as in its patterns of use. This is particularly so, given the results and discussion of the account gathering study.

It will be recalled that pre-diagnostic and diagnostic interpretations of clinical information are described in terms of the clinical problem solver locating the piece or array of information elicited in relation to some segment or section of his own cognitive structure of knowledge and experience. Evidence is presented to substantiate this description. It is also shown that this is an active response to all information (section 11.3) but that information may be differently amenable to interpretation, as the identification of pre-diagnostic and diagnostic interpretations and the need for further clarifying enquiry indicate. It is also indicated (section 11.7) that there is a reciprocal relationship between the subject's activity of assigning structure to clinical information as it is elicited and his reception of further information which may alter those assigned structures. Thus structuring of information is not a single cognitive act but a process in itself of dynamism and change. The discussion in section 4.1.3 substantiates this conclusion. Structuring is more often restructuring than the primary interpretation of information. The account gathering study demonstrates this quite clearly.

It is shown that all groups of subjects make multiple interpretations of the clinical information of both types (section 11.2). Thus information can be structured in many ways by one clinical problem solver. It is also suggested that the problem solver often has the ability to structure the information in many ways other than those he is considering at any one time. In section 11.5 it is shown that restructuring can occur either in the presence or absence of new information and that there are many mechanisms which can give rise to the cognitive operation.

It is clear, then, that both the questionnaire study and the account gathering study substantiate the centrality of the concept of structure in the diagnostic thinking process. Reference to the general psychological literature may enable us to understand this concept and its role more fully. Bart and Smith (1974) provide a useful and relevant discussion of the concept of cognitive structure which, they indicate, has no definition shared by all psychologists. However, their definition of the term is as it is used in the present study: "A cognitive structure reflects the organisation of thought of an organism ... for some interval of time". Such a definition encompasses the changing structures which have been inferred when comparing the results of students and registrars in the questionnaire study, and the three groups of the account gathering study. Such changes in structure have been attributed, largely, to the changing contexts of knowledge acquisition and rehearsal. Bart and Smith's (1974) formulation emanates from a structuralist and logico-algebraic approach to cognitive psychology and so, not surprisingly, defines the two parts of cognitive structure as elements which are the input and output, or content, and processes which are "the representational actions performed on the elements; they are defined in terms of the elements they act on and the elements they produce". Such definitions reflect our differentiation between skills structures and cognitive processes.

Ausubel et al (1978) discuss the role of the cognitive structure in problem solving. This clearly relates to the current context:

"That existing cognitive structure plays a key note in problem solving is evident from the fact that the solution of any given problem involves a re-organisation of the residue of past experience so as to fit the particular requirements of the current problem situation." (p. 571)

This statement surely explains further the results of the questionnaire studies which indicate that the problem solving process develops and changes with clinical practice, either in terms of the skills structure or of cognitive processes or

both. It particularly would appear to give support to the inference made earlier that the contextual aspects of clinical practice, being different from those of medical education, give rise to changes in the diagnostic thinking process or, as it is here put, in the elements⁽¹⁾ and processes of the cognitive structure. Further indication is given of the mechanism of change:

"... cognitive structure is related to problem solving in a repository as well as in a determinative sense. The substantive or methodological product of a problem solving process is incorporated into cognitive structure in accordance with the same principles that are operative in reception learning." (ibid. p. 572)

It would seem reasonable that both elements and processes can be substantive products. Our results surely suggest this conclusion. The endocrinology questionnaire study shows that students and registrars are very different in the extent to which their diagnostic thinking processes rely on cognitive operations not measured by that questionnaire. However, the account gathering study shows quite clearly that in all the categories identified for all research hypotheses, no statistically significant difference is found between students, house officers and registrars. It must, therefore, be concluded that students have in their repertoire of cognitive processes the same operations as registrars but that the extent, manner and relative use of those operations in the diagnostic thinking process differ very greatly between the two groups. Thus, the substantive product incorporated into the cognitive structure in the new context of clinical practice in endocrinology is one of process as well as substance.⁽¹⁾ The results of the neurology questionnaire study indicate that such a substantive product can be in terms of elements in that the difference between students and registrars here is primarily in skills structure not in cognitive processes.

Cross referencing of the findings of the two parallel studies and general psychological theory has yielded a picture of the nature of cognitive structures and of their constructive, mutating quality. The next stage in the discussion, therefore,

Footnote (1) - see end of chapter.

must turn to the mechanism of change itself and so to the nature of the reciprocal relationship between the problem solver's activity of assigning structure to clinical information as it is elicited and his reception of further information which may alter those assigned structures. To these are now added the further dimension of consequent change in the very structures to which the received information is referred and from which assigned structures are selected.

The account gathering study yielded much evidence of a process of multiple interpretation and re-interpretation of the changed or unchanged array of clinical information. This evidence is discussed in detail in sections 11.2 and 11.5. Bruner (1957) discusses the cognitive activity of structuring information in terms of hypothesis generation and testing, which frame of reference has been rejected in the present context. He also uses 'theory construction' and coding systems as explanatory concepts and in his discussion describes the process of structuring and restructuring information:

"If it is a good theory - a good formal or probabilistic coding system - it should permit us to go beyond the present data both retrospectively and prospectively. We go backward - turn around on our own schemata - and order data that before seemed unrelated to each other. Old loose ends now become part of a new pattern". (p. 221)

Bruner defines a coding system as "the person's manner of grouping and relating information about this world, and it is constantly subject to change and re-organisation". It is difficult to appreciate wherein this differs from the elements and processes of the structures which are used as the current explanatory concept and frame of reference.

However, theory construction is a useful concept since Bruner sees it as a creative, inventive process. It is not merely one of referring to a stored structure, but results in new structures to explain the given events and predict their concomitants. Thus existing structures become re-organised in the face of new arrays of information with which to deal. Such

a concept, then, relates the reciprocal assignment and reception activity identified in the account gathering study and the additional developmental dimension of altering stored structures through the process of using them. Bruner also identifies cases of coding processes (or structuring) which could not necessarily result in alterations to stored structures. These are simple placement of the given information in an identity class and learning the probability texture or redundancy of the environment. This latter process, however, would seem to be akin to developing new stored structures such that the problem solver can predict the likely concomitants of a given event.

The concept of theory construction may be seen as related to the classical gestalt theory of productive problem solving which is also described as a creative and inventive process relying heavily upon the problem solver's capacity to restructure the problem in different, progressive ways and to see the inner structure of the situation (Hilgard, 1964). Solutions are therefore based on cognitive re-organisation. Since this is clearly the operation^{here} postulated as a defining characteristic of the diagnostic thinking process, it is worth exploring it further.

Vinacke (1952) points out that productive problem solving is conceived to be a dynamic, fluid process, depending upon a grasp of the inner relations of the problem and an orientation to its structural requirements. A mechanical application of rules and principles is excluded. Past experience is important but what matters is now, and what one recalls and how one applies what is recalled. The concept of productive problem solving originates from the work of Wertheimer (1945).

Although the gestalt context of the concept is not without its important critics (Piaget, 1968) the centrality of the operation of restructuring would appear to be a fundamental cognitive process, regardless of context. It is one of the major processes for which the present account gathering study provides evidence. Maier (1931) suggests that cognitive reorganisation is experienced suddenly and often inexplicably, but that a

change in the perception of one element may organise other elements into a new structure. The results of the account gathering study indicate that sources of restructuring are many and varied (see section 11.5) and may occur either in the absence or presence of new information.

The differences between restructuring as the classical gestalt psychologists meant it in holistic terms and as it is demonstrated in the present study may not only derive from the theoretical framework itself, but also from the nature of the problem under consideration, and the psychological derivation of the restructuring process. A problem of the order "How can solution X be achieved?" where a, b and c are the elements and constraints of the problem, is not of the same type as "What disease, if any, does this person have?" The former is circumscribed, although within broad limits, whereas the latter has no prescribed boundaries or parameters or solution and no specified criteria against which to judge whether or not a solution is acceptable. Barrows (1976) describes the clinical problem as an 'unknown problem'.

Problems exhibiting such characteristics as these have been termed ill defined (Reitman, 1964) as opposed to well defined problems associated with which is some systematic way of deciding when a proposed solution is acceptable. Reitman (1964) considers that most research in the area of problem solving concerns the well defined type and cites, for example, formal bodies of problem solving methods such as linear programming, and information processing approaches towards human problem solving. We may add such classical work as that of Miller, Galanter and Pribram (1960), or a list of component problem solving skills identified from a literature survey for the development of a problem solving test (Speedie et al, 1973) which does not differentiate between ill defined and well defined problems. In addition, we may cite the problems on which the gestaltists mentioned based their theories of productive problem solving.

Duncker's (1945) definition of productive problem solving was as follows:

"The final form of a solution is typically attained by way of mediating phases of the process, of which each one, in retrospect, possesses the character of a solution, and, in prospect, that of a problem".

Here, Duncker is referring to the problem solver's progressive restructuring of the problem and at each stage the suggested solution must be examined for its viability and accepted or rejected according to objective criteria. However, given the ill defined diagnostic problem, structuring and restructuring would seem to serve a different psychological and problem solving purpose.

Reitman's (1964) discussion concerns the transformation of the ill defined problem into one sufficiently well defined to admit of solution; this may be done by addition of appropriate constraints, assumptions, or identifying statements. In information processing terms, this has the effect of reducing or closing the problem space. In the terms of the present discussion, the constraints, assumptions and identifying statements are derived from stored, structured knowledge and have the effect of structuring the information in ways analogous to those identified by the gestaltists but in an entirely different context. Relevant discussion has been presented by Elstein et al (1972) who cite Bartlett's (1958) paradigm of the open system in which the problem solver moves from a known starting point to an unknown terminal point. The processes of structuring and restructuring can relate such an open system to a series of closed ones and test their relative fits. The interpretative processes discussed in sections 11.2, 11.3, 11.4 and 11.5 above may be seen in these terms. However, it must be remembered that, unlike the gestaltist problems, the field of information in the clinical interview changes throughout the problem solving process, thus restructuring may be in the context of an extending field rather than being a process of recentring in a fixed field of information.

The cognitive process of structuring and restructuring arrays of information has been noted by many workers and interpreted within many theoretical frameworks other than those of Bruner and the gestaltists. Ability to structure and restructure has been interpreted in relation to the field dependent/field independent dimension of cognitive style (Shapson, 1977; Witkin et al, 1974). Figure 11.2 for example, may be seen as demonstrating field independence. Such an ability is also defined as distinguishing Piagetian formal thought from prior stages of development (Inhelder and Piaget, 1958; Lunzer, 1968) while the reciprocal relationship between the responsive and deterministic functions of structures and the consequent changes in the diagnostic thinking process (section 11.7 above) may also be discussed within the Piagetian assimilation-accommodation model (Furth, 1968). Relevant extensions of this model (Riegel, 1973; Youniss, 1974) also suggest that problem finding characterises certain forms of mature adult thinking as much as does problem solving (Cropper et al, 1977). The ill defined diagnostic problem could be seen in such terms, although its deliberate and formal nature may invalidate the comparison. However, such discussions of Piagetian theory indicate that the full propositional logic of the stage of formal operations may not always be applied to all problems by those who have the ability (Nixon, 1973). Such a distinction between competence and performance seems to give support to the current preference for a broad psychological, explanatory framework based on concepts of structure and extrapolation rather than the narrower one based on hypothesis generation and testing with its implicit assumptions of cognitive operational method. Despite this, the concepts of structure and extrapolation do not exclude, but rather encompass, the major features of the Piagetian approach. Indeed, the general hypothico-deductive nature of adult thinking is pre-supposed if structure and extrapolation are to make sense. Inhelder and Piaget (1958) state that objectivity pre-supposes a decentring or a continual refocussing of perspective and a differentiation and co-ordination of perspectives and points of view. The results of the account gathering study and the subsequent discussion of struct-

uring and restructuring are in accord with these postulates. Likewise, the current interpretative framework is accepting of the Piagetian definition of a cognitive operation as a type of action which can be carried out either directly in the manipulation of objects, or internally when it is categories or, in the case of formal logic, propositions which are manipulated. An operation is thus a means for mentally transforming data about the real world so that they can be organised and used selectively in the solution of problems. No incompatibility is perceived between this and the explanatory framework of structure and extrapolation which appears best to describe the essence of the diagnostic thinking process.

We may now attempt to summarise the argument and conclusions as follows:

1. The diagnostic thinking process involves the assignment of structure to arrays of clinical information.
2. The structure, and therefore accessibility, of stored knowledge varies with relative experience of medical education and clinical practice.
3. Structuring clinical information is not a single cognitive act, but a reciprocal process of dynamism and change. Cognitive structures have both repository and determinative aspects.
4. Cognitive structures comprise both elements (content) and processes (operations on content). This reflects the differentiation between skills structures and cognitive processes.
5. The diagnostic thinking process yields a substantive product which may be incorporated into the cognitive structure of the thinker, and may be either element or process.
6. The cognitive processes of structuring and restructuring are incorporated into a wide range of theories: coding systems and processes and theory construction (Bruner);

productive problem solving (Wertheimer, Duncker and other gestalt psychologists); the field dependent/independent dimension of cognitive style theorists (Witkin); and Piagetian theory.

7. The process of assigning structure to the array of clinical information renders the ill defined diagnostic problem potentially well defined.

The specific meaning and application of these points are returned to later in this chapter when the three stages (initiation, progress and resolution) of the diagnostic thinking process are considered in a psychological perspective. Meanwhile, the concept of structure having been somewhat clarified we may turn to the second fundamental psychological feature of that process to which the current studies have pointed.

12.1.2 Extrapolation

In section 11.2 above the diagnostic thinking process is described as one of "trying to make sense of the information elicited by referring to other information not present". Thus a pre-diagnostic or diagnostic "interpretation refers to an extrapolated context or array of information of which the information already elicited could be an exemplar if accompanied by certain other items of information". The whole of section 11.4 elaborates on this by discussing the evidence of the account gathering study that clinical problem solvers do work within an extrapolated cognitive context or contexts. The process which others have identified as hypothesis testing by definition requires such reference to features not yet shown to be present, whether these are items of information or relationships.

The process of extrapolation is closely related to that of structuring and together these two operations form the interpretative cognitive activities of the diagnostic problem solver. Indeed, it is reasonable to contend that the operation of extrapolation is contemporaneous with the process of structuring the elicited information and is dependent upon the

appropriateness, completeness and accessibility of the stored structures of information. This is discussed further in section 11.8 above in terms of students' difficulties in retrieving relevant information during the process of the clinical interview.

The importance of the operation of extrapolation is also reflected in the predominant role of the interpretation of symptoms and signs in predicting diagnostic acumen in the questionnaire study for both endocrinology and neurology, although it will be recalled that this importance is somewhat equivocal for endocrinology registrars. Interpretation is clearly normally only possible by reference to stored information. The account gathering study indicates that such interpretation and reference may be more or less satisfactory to the interpreter and, if unsatisfactory, further information is sought to facilitate other interpretations or extrapolation to other contexts (section 11.3). This study also indicates the multiplicity of extrapolated contexts which may arise from an array of clinical information which may or may not alter (sections 11.2 and 11.5). This, again, indicates a close relationship between structure and extrapolation, since location of a range of extrapolated contexts also must imply a similar range of structures assigned to the information array. Section 11.9.2 (figures 11.9 and 11.10) indicates that failure to extrapolate to the appropriate context is one source of error in the diagnostic thinking process and it is suggested that this, in essence, is due to failure of structure. The concept of the forceful feature is discussed and is returned to in later sections of this chapter.

Perhaps the process of extrapolation as discussed here is analogous to the process which others have identified as clinical inference. Hammond (1966) defines the process of inference as "making a judgment about an object or event on the basis of more or less insufficient data". Clinical inference might be of the inductive type, involving reasoning from a particular event to the general case. Engle (1964) conceptualises diagnosis in these terms. As an example of this, Hammond describes the case

where "if one observes a regular relation between symptom 'a' and disease 'A' in one hundred cases, we may reason that all patients who exhibit symptom 'a' have disease 'A'. Deductive inference, on the other hand, refers to a process of relating "phenomena to general principles or events tied by a necessary relationship with those being investigated" (Peel, 1971).

Hammond (1966) considers inductive inference to be most frequently displayed in clinical situations and to be of an imperfect and uncertainty geared type because of the imperfect and uncertain nature of clinical information itself. Cohen (1972) refers to this type of uncertainty as "intrinsic uncertainty", a definition "pertaining to the imprecision, ambiguity or other limitation of the data on which a medical judgment is to be based".

Sarbin et al (1960) consider that the nature of clinical inference varies considerably among persons and that conclusions are not invariant; characteristics of the inferring person influence both the choice of data and the manipulation of terms. The degree of similarity or difference between the stored structures of individuals would surely render this opinion more or less true if inference is taken to mean extrapolation to undemonstrated but possible concomitant features.

It is important to note, that any process of inference, whether inductive or deductive, involves a hypothetical aspect, since characteristics inferred are, by definition, not directly observed and require some form of proof or substantiation. Medawar (1969), in his discussion of intention in scientific thought (as an example of which he cites the diagnostic process), describes scientific reasoning as "a kind of dialogue between the possible and the actual". In other words, it is hypothetico-deductive. It may appear that the process of inference thus described resembles the process of "formulation of possible explanations", described by Peel (1971) as a capability called for when making a judgment. It may reasonably be concluded from this brief discussion that the processes of inference and extrapolation are closely related in respect of the diagnostic thinking process.

We may now turn to the meaning of extrapolation as it is used in the psychological literature, and in contexts other than the present one in order to achieve a more accurate interpretation and understanding of this aspect of the diagnostic thinking process.

The work of Bruner has already been discussed in relation to principles of teaching and learning (Chapter Four) and to the psychological meaning of the concept of structure (section 12.1.1). His work is also relevant to the psychological concept of extrapolation. Its theoretical emphasis has been upon the construction of internal models or generic coding systems by means of which an individual might predict, extrapolate or go beyond the information given (Bruner, 1973). The cognitive activity of extrapolation can, according to Bruner, take a number of different forms. For example:

"Given the presence of a few defining properties or cues, we go beyond them to the inference of identity. Having done so, we infer that the instance so categorised or identified has other properties characteristic of membership in a category ... The act of rendering some given event equivalent to a class of other things, placing it in an identity class, provides then one of the most primitive forms of going beyond information given". (p. 219)

A second form of extrapolation occurs when an individual learns the probability texture of the environment and therefore can predict the likely concomitants of a given event. A third form of extrapolation Bruner identifies as "the learning of certain formal schemata that may be fitted to or may be used to organise arrays of diverse information". This process Bruner calls coding (Bruner, 1973) and it amounts to the methods and rules which an individual may apply to data in order to understand them. One may see this activity as more structural and less content based than the first two forms of extrapolation, in that the same coding process may be applied to entirely different forms or content of information arrays.

The final form of extrapolation which Bruner identifies as
• theory construction enables retrospective and prospective extra-

pulation. This is discussed in section 12.1.1. above.

It is clear that the hypothesis generation and testing model of the diagnostic thinking process adheres very closely only to the first of Bruner's forms of extrapolation; to that which he designates the most primitive form. Given the range and variety of clinical problems, their manifestations, patients' reports, interview structures, and types and structures of knowledge, it seems inappropriate to confine what de Groot (1965) terms "the omnipresence of anticipation" to such an approach which is element dominated. In section 12.1.1 it is pointed out that the cognitive structure is made up of both elements and processes. For example, the mechanism of locating possible appropriate extrapolated contexts or structures is discussed in section 11.9.2 in terms of identification of the forceful feature of an array. It would be reasonable to suggest that such identification concerns some knowledge of the probability texture of the clinical environment such that some items of information have fewer potential meanings or extrapolated contexts than others and so are perceived as more reliable. This is discussed in terms of the selecting operation. As examples, chest pain on exertion or pain in the calves on walking may have fewer potential interpretations than headache or weakness of the legs. Many different processes of extrapolation which Bruner identifies could reasonably be inferred in such a manner from results of the account gathering study. However, the utility and validity of such an exercise is questionable. Likewise, the generalisability of the discussion would be severely limited in the absence of close analysis and identification of the variables of the separate contexts and their characteristics. The present discussion is intended only to demonstrate that extrapolation, like structuring, is not a unitary process. The clinical problem solver has a choice of modes of extrapolation which may sometimes conflict with each other and sometimes operate to the same effect. To characterise this stage of the diagnostic thinking process simply as hypothesis generation is surely to obscure or deny the potential complexity and

range of the cognitive operations which the problem solver may apply.

Consideration of Bruner's work has caused us to conclude that the approach and results of the account gathering process may conceal a rich variety of cognitive function.

Other workers who have contributed to the explanation of the cognitive process of extrapolation or going beyond the information given include Peel (1968, 1971) who distinguishes between describing and explaining:

"Description ... merely involves an account of the phenomenon and a relating of its parts without reference to other ideas, analogies, similarities and antecedent or contiguous circumstances. Explanation ... entails referring the phenomenon causally to previous phenomena and independent generalisations". (1971, p. 26)

Peel considers it "realistic to think of merest description and fullest explanation at the bottom and top end respectively of a scale of ascending power of explanation". But this reservation is presented as a function of the necessary use of language which is itself often conceptual. However, Peel points out that predictions can be made on the basis of either describer or explainer thinking, although prediction from the former are more limited. This reflects our previous conclusion that extrapolation may take different forms at varying levels of complexity. Peel's approach and analysis are very closely associated with the Piagetian model of cognitive development. Such a model, in particular, characterises mature thinking in terms of hypothetico-deductive processes (Inhelder and Piaget, 1958; Lunzer, 1968) which are based on the construction of hypotheses which "entails an abstraction from the attributes of reality as experienced and a reconstruction of reality as the manifestation of a lawful system" (Lunzer, 1968). The most fundamental characteristic of formal thought is the reversal of direction between reality and possibility such that the former becomes merely a special case of the latter (Inhelder and Piaget, 1958).

Thus formal thought is characterised by a process of extrapolation from current information to possible hypothetical explanations. It is particularly salient to the present discussion that such extrapolation is described in terms of a process of cognitive structuring and restructuring (Lunzer, 1968). Indeed, it would be reasonable to infer that even Peel's (1971) describer thinking, which involves selection and relating of parts of a phenomenon, can only occur with recourse to the cognitive structure of the thinker which allows such analysis and construction.

This brief discussion enables us to appreciate the currency of the concept of extrapolation in terms of certain schools of psychological thought and certain psychological theories. It also provides some indication of the close relationship between extrapolation and structure. We return to these concepts, their inter-relationship and possible mechanisms in the diagnostic thinking process in the subsequent sections of this chapter which deal with the three stages of that process: initiation, progress and resolution. Meanwhile, the concept of extrapolation as it applies to our material may be summarised as follows:

1. The account gathering study provides ample evidence that all subjects work within a context or contexts extrapolated from the given information.
2. The interpretative cognitive activity of the diagnostic thinking process may be explained in terms of the two contemporaneous processes of structuring and extrapolation.
3. The importance of the process of extrapolation is reflected in the predominant role of the interpretation of symptoms and signs in predicting diagnostic skill in the endocrinology and neurology questionnaire studies.
4. An extrapolation may be more or less satisfactory to the clinical problem solver.
5. Multiple extrapolations are made by all three groups of subjects in the account gathering study, regardless of whether or not new information is added to the array.

6. Multiple extrapolations imply multiple structurings of the information.
7. Failure to extrapolate to the appropriate context occurs as an error in the diagnostic thinking process.
8. Extrapolation may be seen as analogous to the process of clinical inference.
9. Hypothesis generation and testing is only one of many possible mechanisms of extrapolation, and appears limiting and element dominated and may require prior processes of greater complexity for its operation.
10. Extrapolation is not necessarily a unitary process.

12.2 Stages of the Diagnostic Thinking Process - A Psychological Perspective

This section discusses, in turn, the three stages of the diagnostic thinking process. It must be noted that these have been identified logically rather than psychologically. Therefore, no overall phases of the process are suggested. Instead, it is suggested that the three stages of initiation, progress and resolution should more usefully be applied to the process of interpretation of the changing array of clinical information, and therefore to the processes of structuring and extrapolation. Results of the account gathering study show quite clearly that many interpretations of the clinical information are made throughout the course of the interview and that these are assessed for their appropriateness and dealt with accordingly. Thus, at any point in the diagnostic thinking process, the clinical problem solver may be at the stage of initiation of one interpretation, of progress with another and resolution with yet a third. These three stages, then, are not mutually exclusive and, indeed, the same piece of information may give rise to two stages of thought (for example, a piece of information which causes resolution of one interpretation may also cause initiation of another). Thus, the three identified stages are aspects of the diagnostic thinking process, but are only consecutive stages in the process of one

train of thought based on one interpretation. Since many interpretations are made in concert, it is reasonable to assume that at any one time the diagnostic thinker may display the characteristics of any one, or two or all three stages of thinking and that these may or may not be systematically related depending upon the relationship between interpretations. Thus, the following three sections should answer these questions: What does psychological theory have to contribute to the discussion of research results about (a) how interpretations of clinical information are initiated; (b) how they are cognitively manipulated while still current; (c) how they are resolved or lose currency?

12.2.1 Stage One: Initiation of Interpretations

In this section we may draw together our conclusions concerning the mechanism which gives rise to an initial interpretation of an array or piece of clinical information. We may further relate such conclusions to relevant psychological theory. The importance of such interpretative activity is clearly not to be underestimated. Section 8.3. (above) points out that, with regard to the results of the questionnaire studies, "In the absence of accurate interpretation of symptoms and signs, the primary data, the other processes of diagnostic problem solving will be jeopardised... This principle applies to both students and registrars". Thus the effects of knowledge and selecting and testing diagnostic possibilities appears, from those studies, to be mediated by the interpretation of symptoms and signs. Such a conclusion seems self-evident. Likewise, section 11.7 points out the extremely and consistently active and organising nature of the clinical problem solver's involvement.

Section 11.5 concerns the implications of the account gathering study for the mechanism of cognitive restructuring or reinterpretation. It is stated that the research method is not sufficiently rigorous to allow identification of original or primary interpretations of clinical information. However, although it is the case that the data do not allow such identification in terms of a specific time or place in the interview,

it is reasonable to infer the nature of primary interpretative processes from the conclusions related to mechanisms of re-interpretation. Section 11.12 makes such inferences. It is suggested that the clinical problem solver can only embark upon the interpretative process by means of reference to information not present or, it may be added, to a stored structure of similar information where all necessary information is simultaneously present but requires structuring in order to become meaningful. For example, the simultaneous presentation in a female patient of a history of an apparently serious symptom, such as paralysis, given in an unconcerned manner would admit of the diagnostic interpretation of hysteria (*la belle indifférence*). However, the question of how such reference is made is the central one for the present discussion.

The results of the account gathering study show that either pre-diagnostic or diagnostic interpretations, or both, may be made as a primary response to clinical information and that the selection of either is primarily a function of the information elicited rather than the propensities of the thinker. It is suggested that information is not stored in the problem solver's memory simply as a series of discrete arrays each of which represents a pre-diagnostic or diagnostic interpretation, but rather that information is also stored such that the problem solver may recognise a forceful feature and make his connections from there. This mechanism for identification of the forceful feature of an array, we may term the selecting operation. Such a forceful feature may be the item or grouped items of information with the greatest payoff in terms of facilitating productive interpretations or structuring responses. Further, such identification of the valuable information may change with experience as the process is re-inforced or not. For example, given a history of breathlessness and swelling of the hands and ankles, the less experienced problem solver may extrapolate to cardiac failure having identified breathlessness or ankle swelling (or both) as the forceful feature. However, the more experienced clinician may identify the

swelling of the hands as the forceful feature and extrapolate to a diagnostic interpretation of systematic cirrhosis. It will be recalled that Thyne's (1966) concept of the forceful feature is discussed in section 4.1.2 above and we may pursue its relevance further in our discussion of the origin of interpretative response.

In section 11.9.2 it is suggested that structuring or interpretation in the diagnostic thinking process almost invariably requires the problem solver to select significant items of information from among the array of relevant and irrelevant items and to inter-relate them appropriately according to some criterion or criteria. Failure to identify the true forceful feature and structure the information accordingly may be one source of error. It is pointed out that information can only be interpreted and structured according to the limitations of the subject's own information store in both its structure and its content. These features have, in turn, direct implications for accessibility of the store. It is such ease and appropriateness of access which must eventually determine the efficiency and effectiveness of initial interpretations and, by implication, of the entire diagnostic thinking process. We suggest that access to the stored structure is achieved by means of some preliminary structure which enables identification of forceful features from among an array and extrapolation from there to possible embedding contexts.

To understand the nature of a forceful feature, we may remind ourselves of subject R. 02, whose response to vertigo or weakness was to "cringe and crumble", whereas "discomfort in the belly" was "a gem". R. 02 demonstrates very clearly that different pieces of information are differently amenable to interpretation (see 11.5). That is to say, they have different forcefulness. We may also refer to the discussion of the set effect (section 11.9.1) which illustrates quite clearly that sometimes information may be designated as a forceful feature when it is not, in fact, the cue of the array (see 4.1.2). The cue, here, is the feature or features which enable extrapolation or access to an appropriate (but not

necessarily correct) stored structure of information which is the pre-diagnostic or diagnostic interpretation. In the case of R. O2 it may be inferred that vertigo and weakness do not constitute the forceful feature of any or many arrays of information. They are not cues from which he can readily extrapolate to possible embedding contexts even though, without doubt, such information is located in a number of structures stored in his memory. But these are not the pieces of information which allow either easy access to those structures or access to structures of sufficient definition and use value. In the case of the set effect, it is clear that an incorrect forceful feature or features may sometimes be identified which, in turn, directs the problem solver to an inappropriate embedding context where his thinking remains locked unless a reappraisal of the array of information and re-identification of the forceful feature occurs. Thyne (1966) discusses the correct identification of the forceful feature (that is, identification of the cue) and its relationship to understanding. His context is with regard to learning rather than problem solving, but the same principles apply to both situations:

"If understanding is to take place, not only must the learner see the cue, and see it within the material to be understood, but he must already be familiar with the cue - he must already know it. The cue he has to see in this particular material is something he will recognise once he does see it". (p. 132)

In these terms, then, we may interpret seeing the cue as identifying the forceful feature, and understanding as extrapolating to a possible embedding context. Such extrapolation Thyne (1966) also conceptualises in terms of the forceful feature and the cue:

"If the forceful feature of the present situation is indeed the cue of some established instance of learning, of necessity the learner will respond to the present situation as he responds in that established instance, for the reason that the present case is a case of that very instance". (p. 224)

Again, we must transpose the framework of learning into that of problem solving. In essence, Thyne is saying that one relates new and stored arrays of information in terms of their forceful features or cues. With regard to the diagnostic thinking process, it is being suggested similarly that in the first instance a forceful feature is identified and from there the problem solver can extrapolate to the arrays of information which are structured by means of their forceful features. Such a mechanism might possibly explain the finding that some subjects respond to information by judging that they need further clarification before extrapolating to an embedding context in which they are happy to work (section 11.3). Thus features of the array are identifiable, but none of them attains a sufficient threshold of force to be useful in an effective or efficient or otherwise satisfactory problem solving process.

One related inference must be made with regard to this discussion of the relationship between the forceful feature and extrapolation. It is noted in 11.3 that a subject, if presented with any piece of clinical information, could almost certainly extrapolate from it to many contexts, but to do so in some instances would be grossly inefficient, therefore further clarification is demanded. However, this very inference must in turn imply that stored structures of information can be located not only via the forceful feature or cue of the array, but via other features which are not either the forceful one or the cue. It would seem reasonable to suggest that the latter manner of access is likely to be less efficient and effective than the former. For such reasons, we have inferred that forcefulness is not a relative but an absolute value for any one thinker, and that a sufficient threshold of force must be attained before any item of information is useful for extrapolation. However, this further illumination of the role of the forceful feature does not explain the mechanism whereby that feature (whether it be one or a number of items of information) is identified from among the array of inform-

ation presented. As Thyne (1966) indicates, the learner (or problem solver) must "see the cue, and see it within the material to be understood". However, he offers no explanation of how this perception occurs. We must, therefore, turn to other theorists.

It has already been suggested that flexibility in the origin of interpretations may be related to some preliminary structure which enables identification of forceful features, or features with the greatest interpretative payoff, within an array and extrapolation from there to possible embedding contexts. For the moment this must remain a hypothetical construct only, but one which is, nonetheless, amenable to objective testing. Before pursuing such a concept further, however, it must be noted that it is not being put forward as a necessary stage in the diagnostic thinking process but an available one where an array of information makes some selective assessment and possible identification of a subset of information necessary. For that reason we may refer to the suggested mechanism as the "selecting operation".

Crutchfield (1972) refers to the first stage of problem solving as problem discovery and considers that:

"... to attack the problem effectively requires that it then be formulated in a meaningful way by the individual; he must state the problem to himself in a form which points up the crucial issue to be solved and which avoids a misleading set". (p. 193)

Thus the problem solver must, in essence, organise the field of information appropriately and must select from among the information available that which is most productive. To discover the problem in a clinical setting must be to identify what is actually wrong or is not as it ought to be (that is, to discover the actual symptoms and signs) and to follow on from that to identify why this is so and how it may be corrected. Ausubel et al (1978), in a pedagogic context, use the concept of advance organisers. These constitute a strategy for:

"... deliberately manipulating cognitive structure so as to enhance proactive facilitation and to minimise proactive interference ... These organisers are normally introduced in advance of the learning material itself and are used to facilitate establishing a meaningful learning set. Advance organisers help the learner to recognise that elements of new learning materials can be meaningfully learned by relating them to specifically relevant aspects of the existing cognitive structure". (pp 170 - 171)

If we transpose this into the problem solving context it is clear that a process similar to our selecting operation is being discussed. Enhancement of proactive facilitation may surely be achieved by identification of appropriate pre-diagnostic or diagnostic interpretations to be measured against reality. This may be achieved directly by immediately identifying the congruence between given information, as it is presented, and information stored in memory. Or it may be achieved indirectly by firstly selecting certain information from among the given array and extrapolating from there to the stored structures. Minimisation of proactive interference may be achieved by ensuring that the response to the given array of information does not yield extrapolation to inappropriate contexts. Thus proactive interference in the diagnostic problem solving process may occur if a given array of information is accepted without selection of an appropriate subset and extrapolation occurs on the basis of objectively unproductive information. In other words, proactive interference may occur where a forceful feature is identified only by default and where this is not actually the cue. The hypothesised selecting operation allows active identification of the forceful feature by the problem solver, rather than passive acceptance of force on the basis of order or emphasis of information as it is given. It must be noted, however, that the concept of a selecting operation refers to a cognitive process. The content of that process need not, necessarily, be correct or appropriate in any given case. Thus a selecting operation may yield a forceful feature which is not a cue. However, it has already been pointed out (section 11.9 above) that any aspect, process or strategy of clinical problem solv-

ing is inherently a potential source of error depending upon the knowledge and experience of the problem solver as well as efficiency of implementation of the process itself. Ekwo and Loening-Baucke (1979) using videotape and simulated patients, show very clearly for example, the inhibiting effect which lack of knowledge alone can have in the diagnostic process.

Ausubel et al (1978) describe advance organisers as being used to facilitate establishment of a meaningful learning set by helping the learner to recognise that elements of the new array can be related to specific relevant aspects of the existing cognitive structure. This is precisely the function proposed for the selecting operation in enabling identification of the forceful feature from among the array and consequent efficient extrapolation to appropriate embedding contexts. This function of selection from among the array of information and the opposing concept of acceptance of the relative force assigned to items of information by the order or emphasis with which it is given leads us to consideration of the concepts of field independence and dependence.

Dickstein (1968), in discussing field independence, stresses the capacity to differentiate complex stimulus fields and to deploy attention selectively toward those aspects of the field that are task relevant while ignoring those aspects of the field that are task irrelevant. The selecting operation is suggested as a mechanism which enables such selective attention in the first instance, the relevant task being efficient extrapolation to appropriate embedding contexts. Thus task irrelevant aspects may be so only ephemeral and become task relevant as the problem solving process continues. Witkin et al's (1974) less qualified definition of field independence as "the ready ability to overcome an embedding context and to experience items as discrete from the field in which they are contained" seems to be more directly generalisable to the present context. The concepts of field dependence and independence, of course, initially arose out of research in visual perception. However, Witkin et al (1974) maintain that the

dimension "pervades the individual's perceptual, intellectual, emotional, motivational, defensive and social operations". Field independence Witkin et al (1974) term "analytic field approach". Such a tendency to experience analytically they, in turn, relate to the ability to structure experience.

Other workers have also commented upon the ability to select special features from an array of information. Bartlett (1958) for example, comments upon this phenomenon in relation to the same initial stage of a thinking process as we are here considering. His context, however, was of sectional map reading:

"As the thinker takes up his search in the more open type of system, in the early stages directional features, belonging to the structure, or to an assigned structure, of the system in which he is working, but themselves of a general character, are predominantly effective".

Bartlett's directional features may be interpreted as analogous to Thyne's (1966) forceful features. Abercrombie (1960) calls these clues. Neither Bartlett (1958) nor Thyne (1966) nor Abercrombie (1960) indicate the mechanism whereby these features are located. The concept of a selecting operation is, however, advanced in the present discussion to explain the initiation of the diagnostic thinking process. It will be noted that Bartlett (1958) is considering an open system of which the diagnostic problem is an example and that he identifies the directional features as belonging to either the actual or assigned structure of the system in which the thinker is working. This latter point, in particular, is interesting since it may assist the further definition of the hypothesised selecting operation and its consequences. Rather than belonging to either the actual or an assigned structure of the system or, as in the present case, the array of information with which the thinker is working, our own interpretation of the initiation of the diagnostic thinking process would cause us to conclude that the forceful or directional feature must belong to both the actual and the assigned structures, which may or may not differ. The sel-

ecting operation enables identification of that feature in the array of given information and location of extrapolated, embedding contexts or structures which also display the same forceful feature. In addition, they are effective to the extent that they allow the thinker to work within those systems. As we have seen, on some occasions, no piece of information is accepted as a satisfactory forceful or directional feature in those terms (section 11.3 above). Bartlett considers that, as thinking progresses, these features become less independently important as empirical features and then clusters of features gain predominance. Although Bartlett's sectional map reading context was very particular, the parallels between his conclusions and the present discussion are clear. Once the directional or forceful feature is identified, extrapolated or embedding contexts may also be located and specific features sought in the presenting array. It is worth noting, however, that our evidence of multiple interpretations (section 11.2) may lead us to infer that more than one forceful feature may be identified in any one array of information, and that the same item may be the forceful feature of many different structures or arrays of information.

Before leaving the discussion of the initiation of interpretations we may further relate the selecting operation to other theoretical constructs for their further explanatory power.

To begin with, we may usefully return to our consideration of models based on Piagetian formal operations. Lunzer (1968) points out that the solution of a problem in terms of formal reasoning depends upon the analysis and reconstruction of a system perceived in a way that permits its consideration as a determinate system. This process entails abstraction of relevant portions of the situation itself, abstraction from the thinker's potential actions to bring out a possible order, and the construction of a closed system on the basis of a mental model of the events conceived. Again, the analogy between this description and the processes of the selecting operation and extrapolation to a possible embedding context is clear. The role of the selecting operation is made even

more pertinent by Lunzer's (1968) pointing out that an objective stimulus complex may be presented to a thinker, but that his schemata (or structure or selecting operations) actually determine what will function as a stimulus for him, what he will attend to and what he will retrieve from memory in response. Cohen (1972) also makes the point that information transmitted is not necessarily the information received, although he is considering communication from doctors to patients which is the reverse of the present situation. Thus, the example given in section 11.9.3 above of information being either appropriately or inappropriately designated as irrelevant, may be seen in these terms. Such a selective function of consciousness is not a recently recognised phenomenon, as Miller's (1962) reference to the 1890 work of William James reveals.

Our reference to psychological theory, then, has advanced our understanding of the possible mechanisms operating at the stage of initiation of interpretations of clinical information. Many of our conclusions can be only reasonable inferences and must remain to be tested. However, we may summarise the argument as follows: ,

1. Interpretation of clinical information is a sine qua non of the diagnostic thinking process, mediating between information elicited and diagnosis made.
2. Interpretation of clinical information can only be made either by reference to other stored knowledge which is not present or by seeking a stored structure of knowledge which will facilitate meaningful structuring of the information elicited without reference to other knowledge.
3. Information may be stored in memory in terms of pre-diagnostic and diagnostic interpretations and in terms of a comparative selecting operation which enables identification of the forceful feature of an array.
4. The forceful feature or features identified by the selecting operation will have the greatest payoff in terms of facilitating interpretations or structuring responses. by means of extrapolation to stored contexts.

5. Efficient and effective structuring or interpretation in the diagnostic thinking process requires the thinker to select significant items of information from among the array and extrapolate from there to possible embedding contexts.
6. Different pieces of information have different interpretative value or forcefulness.
7. Items of information may be inappropriately identified as forceful features if they are not actually the cue (characterising feature or 'correct' forceful feature) of the array.
8. The forceful feature of an array is that information which facilitates extrapolation or access to an appropriate (but not necessarily correct) stored structure of information which is the pre-diagnostic or diagnostic interpretation. Thus, elicited and stored arrays of information are related in terms of their forceful features.
9. It is possible to extrapolate from or interpret any piece of clinical information, whether or not it is identified as a forceful feature, but only with greater or lesser degrees of efficiency, effectiveness, facility or personal satisfaction. In some cases the thinker is unwilling to work with any of the extrapolations which he can make. Thus, items of information must attain a satisfactory threshold of force before being perceived as useful in the efficient or effective problem solving thinking process.
10. Forcefulness of a feature is probably not a relative, but an absolute value for any one thinker since evidence suggests that a sufficient threshold of force must be attained before any item of information is useful for satisfactory extrapolation.
11. The concept of the selecting operation is suggested not as a necessary, but rather an available stage in the diagnostic thinking process where the characteristics of the presented or elicited array of information make it necessary.

12. The selecting operation allows active identification of the forceful feature by the problem solver, rather than passive acceptance of force on the basis of order or emphasis of information as it is given.
13. The selecting operation should facilitate field independence and application of selective attention.
14. The forceful feature must belong to both the actual and the assigned structures.
15. Multiple interpretations of information allow the inference that more than one forceful feature may be identified in any one array of information and the same item or items may be the forceful feature of many extrapolated structures or arrays of information.
16. The concept of the selecting operation is congruent with the characteristics of formal reasoning.

12.2.2 Stage Two: Progress of the Diagnostic Thinking Process

This section concerns the cognitive operations applied to the interpretations of the array of clinical information after their initiation and prior to their resolution (or final selection). This stage has two defining characteristics: restructuring and assessment of interpretations by working from the extrapolated context. The discussion is presented in general terms, not noting differences between endocrinology and neurology, or between students, house officers and registrars. This approach can be adopted since such differences are largely quantitative and not qualitative (see section 12.3 below). Thus the discussion applies to a greater or lesser degree to all subjects in both specialities. The differences of degree which do exist between subjects and specialities, as well as their similarities, are discussed below (section 12.3) against the background of the present discussion.

One of the major features of the second stage of the diagnostic thinking process to emerge from the account gathering study is the restructuring or re-interpretation of clinical information.

Figure 11.2 gives an illustration of this phenomenon. Section 11.2 shows that many interpretations may be made of the same information array or of subsections of that array or inter-sections of a slightly changing array. It is pointed out that these different interpretations, logically, can arise in two ways. Firstly, the same array can be structured in different ways (for example, if different elements can be identified as the forceful feature, the other items of information may arrange themselves around that feature in different ways). We must remember, however, that the forceful feature is a psychological construct and in many instances may only be identified by psychological testing and inference. Secondly, the thinker can extrapolate to a variety of contexts from the information given. The latter would seem to be the more usual process and is compared to Peel's (1971) explainer thinking. It is suggested in 11.2 that multiple hypothesis formation is less likely to be for reasons of economy of cognition than a simple function of the multiplicity of potential extrapolated contexts in which the given information can be embedded. It is also suggested that the thinker has the ability to structure the information and extrapolate from it in many ways other than those which he considers at any one time. This gives added weight to our conclusion (9) in the previous section that some extrapolations are more satisfactory to the thinker than others.

It has been shown that progressive specification is not a characteristic of restructuring. Either pre-diagnostic or diagnostic interpretations may be made at any point during the diagnostic thinking process and in any order. This must largely be a function of the nature of the information elicited. It is worth remembering that the clinical information itself is an interactive variable in the diagnostic thinking process.

Section 11.5 shows that restructuring of the array can occur either in the presence or absence of additional clinical information. The re-interpretative cognitive activities of every subject of the account gathering study may itself be interpreted in terms of the usually positive qualities

of flexibility of thought, lack of tendency to rapid closure and field independence. However, it is pointed out that neither form of re-interpretation, nor re-interpretation itself, necessarily has any intrinsic merit in the solution of any one clinical problem.

Having thus reminded ourselves of the major salient features of restructuring or re-interpretation of the clinical information during the diagnostic thinking process and clinical interview, we may consider their psychological interpretation. In section 12.1.1 Bruner's (1957) view of structuring and restructuring is given, but other workers have also presented theories to account for this general cognitive process. Duncker (1945) emphasises that the process of problem solution can be described either as development of the solution or as development of the problem. His definition of productive problem solving, which espouses the latter approach, has already been cited (section 12.1.1) but it is worth quoting him further, particularly since his ideas of phase succession are echoed very closely by de Groot (1965). He is here considering the case where a problem solver must try a completely new approach to the problem:

"... a transition to phases in another line takes place typically when some tentative solution does not satisfy, or when one makes no further progress in a given direction. Another solution, more or less clearly defined, is then looked for ... In the transition to phases in another line, the thought process may range more or less widely. Every such transition involves a return to an earlier phase of the problem; an earlier task is set anew; a new branching off from an old point in the family tree occurs. Sometimes a S returns to the original setting of the problem, sometimes just to the immediately preceding phase".

De Groot (1965) adds to this the concept of successive transitional phases through which "we assume that there is an undercurrent, a practically continuous development of the main problem". Although we may see similarities between the approaches of Duncker and de Groot, they, nonetheless, are not satisfactory descriptions of the clinical problem solving operation of restructuring. This would seem to be due to the

inappropriateness of the concept of phase succession and the associated assumption of productive problem solving as progressive "sharpening and specialising of the problem setting" (Duncker, 1945). De Groot's (1965) related transitional phases are seen as the process of integration of results already achieved which yields problem transformations. Although it is quite possible that a search of the accounts of the present study might show examples of such phase succession, it is nonetheless not a satisfactory general description of the diagnostic thinking process which does not progress through stages of increasing concreteness of specificity, but rather is characterised by an increasing array of information which is either newly structured from time to time or which is congruent with existing interpretations. In addition, such a process involves recourse to learned knowledge, stored and structured in memory. It is therefore unlike the problems upon which Duncker based his theory (for example, the X ray problem or the abcabc division problem) or de Groot's (1965) games of chess. It may well be that the necessary role of a body of knowledge in the diagnostic thinking process is a major characteristic which differentiates it from many of the otherwise relevant theories of problem solving. The potential simultaneity of extrapolation from the given information to many different embedding contexts either of the pre-diagnostic or diagnostic type precludes phase succession of the type described by Duncker and de Groot. Likewise, Duncker's transition to phases in another line is inappropriate if it necessarily involves a return to an earlier phase of the problem rather than a simple re-interpretation of the current position. Despite this, however, the conceptual approach of development of the problem rather than the solution is useful. In addition, we have shown (section 11.3) that subjects do, as Duncker (1945) notes, seek other possible interpretations when "some tentative solution" (or interpretation) "does not satisfy". In these terms, development of the problem implies transforming it from ill defined to well defined by means of extrapolation to a possible series of embedding contexts.

The possibility of explaining restructuring or re-interpreting in terms of theories of field dependence and independence is mentioned earlier. Shapson (1977), in these terms, suggests that "for effective coding (or recoding), a subject must break down the stimulus array into all its component parts". For example, headache, visual deterioration, lethargy and increased frequency of micturition may be structured as visual problems (headache plus visual deterioration) and urinary problems (frequency of micturition). At the same time, lethargy may be taken as indicating general ill health. The pre-diagnostic interpretations of visual and urinary problems accompanied by general ill health may facilitate extrapolation to the diagnostic interpretation of multiple sclerosis. Upon re-interpretation, the four initial symptoms may facilitate direct extrapolation to a context of pituitary tumour with concomitant hypopituitarism and diabetes insipidus. Thus, the ability to dissociate and relate separate items of information is an important one and is a defining feature of field independence. We shall see later that this ability is important not only in manipulating information as it is elicited, but also in restructuring information which is already interpreted. Inappropriate dependence upon the field of the extrapolated embedding context may result in the error of the set effect, just as failure to overcome the embedding context of information as it is elicited (Witkin, et al, 1974) may result in failure to extrapolate to the correct context in the initial instance (see 12.2.1). Lunzer (1968), in a formal reasoning context, perhaps surprisingly, presents restructuring in problem solving in a similar way, suggesting that the solution of problems entails the construction and reconstruction of objects and events out of their abstracted constituents. The concept of abstraction may clearly be related to the concept of overcoming an embedding context.

Perception of separate constituent elements or items of information, however, would seem to be only a necessary but not a sufficient condition of the diagnostic thinking process.

It will be recalled that the selecting operation is hypothesised as a mechanism preliminary to extrapolation. Thus, the forceful feature (or features) may be identified and extrapolated contexts located.

We must now extend this theory to the stage of restructuring. Thyne (1966) describes this process as a shift of force. He suggests that "when I misunderstand something I see not the cue, but some other feature ... the change-over from misunderstanding to understanding entails the cue's appearing ... in place of some other feature ... And the shift from seeing some irrelevant feature to seeing the cue, is in effect a shift of force to the cue". Although Thyne's theory is shown to be thoroughly relevant to the stage of initiation of interpretations (section 12.2.1) this description of restructuring calls it into question with regard to the second stage of the diagnostic thinking process. Thyne's description would only apply if any one piece of information could be the cue of only one structure or interpretation. However, a symptom such as chest pain or diarrhoea can be the cue of many interpretations. Thus the correct forceful feature of an array of elicited information may be identified and a range of interpretations made, only one of which is actually correct. It is also logically true that the same item of information may be the cue of many of these interpretations. Therefore, understanding (that is, diagnosis) does not necessarily arise out of either identifying the cue of the elicited array or out of extrapolating to stored structures of which that item is also the cue. The stage of progress must also involve checking to determine whether or not the other features of the stored array are also present in the elicited history. In summary, identification of the forceful feature facilitates interpretation by structuring and extrapolation. It does not establish the actual validity of the interpretation made. Restructuring can occur by means of a shift of force to a new feature of the elicited array, or by means of extrapolation to multiple contexts on the basis of the same identified forceful feature. Either type of restructuring may occur in

either the presence or absence of new information being added to the array (see section 11.5).

In the discussion of restructuring, the failure to restructure when such an operation would have been appropriate is also mentioned. This phenomenon is discussed in section 11.9.1 as the set effect. As this phenomenon manifests itself in the diagnostic thinking process, it is defined in 11.9.1 as "a tendency to rationalise, ignore or find explanation for incongruent information so that the array elicited fits the identified extrapolated context". The phenomenon, therefore, is not necessarily a failure to make other interpretations, but rather a tendency to favour one. This is also Elstein et al's (1978) interpretation. Figures 11.6, 11.7 and 11.8 give examples of the set effect in a student, a house officer and a registrar. Although a range of cognitive processes or tactics is demonstrated which enables the subject to maintain the inappropriate set, the generalisable conclusion is drawn that the clinical information elicited is not evaluated in its own right, but rather exclusively in relation to expected findings suggested by the extrapolated context which has become the set. In section 12.2.1 it is suggested that an inappropriate forceful feature may direct the problem solver to an inappropriate interpretation or extrapolated context where his thinking may remain locked unless the information is reappraised and the correct forceful feature identified. However, our argument has since developed to add that identification of the cue may still allow extrapolation to inappropriate contexts. The question for psychological theory to illuminate must concern why such a phenomenon occurs.

With regard to the process of perception Bruner (1951) suggests that the concept of an hypothesis is best likened to such a term as set and that it is "a highly generalised state of readiness to respond selectively to classes of events in the environment". He suggests that the strength of an hypothesis increases with greater frequency of past confirmation, with smaller numbers of competing hypotheses, with a larger system

of supporting hypotheses and beliefs, with greater personal or motivational involvement and with greater agreement of the hypotheses of others. While any of these cases may be true for a given instance, the model lacks apparent generalisability, and relies on individual differences and tendencies for its explanation, as does that of Duncker (1945). Luchins' work (quoted in Manis, 1968) cites the reinforcing effects of previous experience in fixing methodological approaches to problem solving, while Wason (1968) considers "dogmatic thinking and the refusal to entertain the possibility of alternatives". Peel (1971) mentions the tendency to "try to hold on" to a view as a characteristic of both sophisticated and immature thinkers and observes that "new explanations may be obstructed by the directive and selective influence of the accepted theory upon the subsequent collection of observations and results". He adds that neglect of contrary evidence may be either conscious or unconscious.

Shapson (1977) suggests that field independent individuals can overcome an embedding context and experience items as discrete from their surroundings. The phenomenon of set seems related to this, in that the subject appears to be dependent upon the psychological or hypothetical field created by his own extrapolation and interpretation and unable to extrapolate the other embedding contexts. Witkin et al (1974) in field dependency terms, suggest that the set breaking process may be conceptualised in terms of the ability to overcome embeddedness. For a set to be broken, the elements must be considered apart from their previously adopted organisation and arranged into a new organisation. Witkin et al (1974) therefore expect subjects with an analytical field approach to display a greater capacity for set breaking than those who have a global field approach. Ohnmacht (1966) provides evidence in support of this expectation, demonstrating that relatively field independent subjects deal with both reversal and non-reversal shifts in card sorting tasks better than do relatively field dependent subjects.

Our discussion of the concept of set in psychological theory does not seem to provide any great degree of explanation of the phenomenon as it is manifested in the diagnostic thinking

processes of the subjects of the present account gathering study. In essence, the theories cited either state that the phenomenon occurs in a variety of contexts and offer particular theoretical frameworks for the description, or explain the phenomenon in terms of particular individual differences and tendencies. We therefore must leave this particular error of the diagnostic thinking process without proper explanation. We have shown that the set effect does not result from failure to restructure information, therefore explanations in terms of field dependency or fixedness are inappropriate. Instead, we may only say that the phenomenon is characterised by a tendency to favour a certain interpretation of information and to rationalise or ignore contrary evidence. It is not necessarily a failure of structuring and interpretation or extrapolation, but may be a failure to test adequately the congruence of interpretation and reality. Thus an inability, for whatever reason, to differentiate expectation and observation is evident. The present study does not allow identification of the conditions of the error. It seems reasonable to suggest that the uncertain and incomplete nature of clinical information and the concomitant necessity for the clinical problem solver to make judgments rather than provide proofs, make an ideal environment in which the error of set may occur. This is discussed further in 12.2.3.

The account gathering study indicates other features which may occur during the second stage of the diagnostic thinking process, but these, like the set effect, are not characterising qualities but variations on the central theme of restructuring and assessment of interpretations by working from the extrapolated context. For this reason we find that psychological theories have little interpretative value here, since the conditions which give rise to these variations are not in any way defined or indicated. We therefore cannot add to the explanations provided in Chapter Eleven. Many of the variations identified take the form of specific errors. For example, failure to make either specific or general enquiry identified in retrospect by the subject is attributed to

retrieval difficulties related to storage structure (section 11.8). Failure to make the correct interpretation where there is no inadequacy of knowledge (section 11.9.2) may also be related to accessibility of the stored information. Our earlier discussion of the forceful feature, the cue and their identification in both the elicited and stored arrays of information adds considerable explanation to this inefficiency of the diagnostic thinking process. The error of inappropriate designation of irrelevance (section 11.9.3) is attributed either to failure of structuring and extrapolation or failure of immediate knowledge. Logically it would seem that all such errors should arise from either inadequate knowledge or inadequate structuring of stored or elicited knowledge. These features might give rise to failure to identify the cue, extrapolation to inappropriate contexts, failure of extrapolation or failure to test the appropriateness of extrapolation. It does not seem possible to make a more generalisable statement, since it would appear that, in adult thinking, systematic errors do not occur. Instead, where errors are identified, as here, they are particular to specific problems or occasions and not even necessarily to specific individuals. Since their conditions of occurrence are not identified, they are also unpredictable. A more productive way of considering them, therefore, may be to identify the conditions which preclude their occurrence rather than encourage it. Such conditions are identified in section 11.9.1.

Having thus considered restructuring and its related processes, we may now move to the second of the two features identified earlier as characterising the second stage of the diagnostic thinking process. This is the assessment of interpretations by working from the extrapolated context which is analogous to the stage of hypothesis testing mentioned by Elstein et al (1978) and central to the work of Barrows et al (1978). We are not here concerned with the manner in which interpretations are finally evaluated, confirmed or rejected, but rather with the effects on the information gathering process of having made the interpretation in the first place.

Section 11.7 provides a discussion of the determinants of the interview structure. It is found that the course of the interview may be determined by the flow of information presented (called patient determined), by the subject's requirement to test his interpretations (called subject determined) or by the logical structure of the standard interview (called logically determined). It is noted that, ultimately, the structure is determined by the interviewer, who will choose to follow one of the three identified courses. The major finding and conclusion is that the problem solver has an overriding need to control the flow of information and that he actively does this. However, this active, determining role is tempered by the reciprocal activity of responding to the information elicited which is unpredictable in its content and so may cause restructuring or some other additional, responsive cognitive activity in the problem solver. Thus information collection, based on the expectations which arise out of the extrapolated context (see 11.4) is a further characteristic of the second stage, but it is qualified by the subject's responsive activity of reinterpretation. It is argued above (sections 11.7 and 11.8) that a certain threshold of data acquisition is a necessary, but not sufficient condition of the successful diagnostic thinking process. This is echoed by the homogeneity of registrars' skills structures which are demonstrated in the questionnaire study (see section 8.6) and the primacy of the predictive power of scores on interpretation of symptoms and signs (see section 8.3).

In conclusion, then, we may suggest that the second stage of progress of the diagnostic thinking process is characterised by the cognitive operations of restructuring and assessment of interpretations by working from the extrapolated context. These operations presuppose and necessitate active, determinative involvement of the problem solver in controlling the flow of information yet, simultaneously, responsive analysis of the areas of content which cannot be controlled. Restructuring and assessment are both, therefore, determinative and responsive cognitive operations. It is contended that all

other operations of the second stage are variants of these central characteristics. Before considering the final stage of the diagnostic thinking process, we may summarise the argument of this section as follows:

1. Multiple interpretations of the same or the changing array of clinical information may be made. Therefore, restructuring and re-interpretation by extrapolation to a variety of embedding contexts are characteristic of the second stage of the diagnostic thinking process.
2. Progressive specification is not a necessary characteristic of restructuring. The pre-diagnostic or diagnostic nature of re-interpretation is largely a function of the nature of the information elicited.
3. Restructuring is the first characteristic of the second stage of the diagnostic thinking process. It has, per se, no intrinsic merit in the diagnostic thinking process in all circumstances.
4. Neither Duncker's (1945) concept of phase succession, nor de Groot's (1965), concept of successive transitional phases provides a satisfactory description of the diagnostic thinking process of restructuring because of their associated assumptions of progressive orderliness and specificity. This is inconsistent with the changing and increasing array of clinical information and the potential simultaneity of extrapolation to different embedding contexts which precludes phase succession.
5. The necessary role of a learned body of knowledge differentiates the diagnostic thinking process from many other forms and theories of problem solving.
6. Restructuring may be explained in terms of field independence and the associated ability to dissociate and relate separate items of information.
7. The concept of restructuring is recognised in a formal reasoning context in terms of abstracted and reconstructed constituents.

8. Perception of separate constituent elements is a necessary but not sufficient condition of the diagnostic thinking process.
9. Restructuring in the diagnostic thinking process cannot be interpreted as a shift of force from an irrelevant feature to the cue, since any one feature may be the cue of many extrapolated contexts.
10. Understanding (or diagnosis) does not necessarily arise out of either identifying the cue of an elicited array or out of extrapolating to stored structures of which that item is also the cue. Diagnosis arises out of establishing the congruence of elicited and extrapolated information.
11. Restructuring can occur either by means of a shift of force to a new feature of the elicited array or by means of extrapolation to multiple contexts on the basis of the same forceful feature. Either type may occur in the presence or absence of new information.
12. The set effect in the diagnostic thinking process is a tendency to rationalise, ignore or find explanation for incongruent information, so that the array elicited fits an identified, favoured extrapolated context. Thus clinical information is not evaluated in its own right, but only in relation to the expected, extrapolated findings of the set.
13. The inappropriate set may be extrapolated from either the cue or an inappropriate forceful feature.
14. Bruner compares hypothesis and set. Each is characterised by selective response to classes of events in the environment.
15. The work of Duncker, Luchins, Wason and Peel in reference to the set effect is cited. It is concluded that these descriptive and specific approaches preclude generalisability to the current context.
16. The set effect may be interpreted as resulting from dependence upon the psychological or hypothetical field

created by the thinker's own interpretative activity, rather than dependence on the field as elicited. Set breaking involves the capacity to overcome embeddedness.

17. Psychological theory seems to offer no useful explanation of the set effect in the diagnostic thinking process since it is not necessarily a failure of structuring and interpretation or extrapolation. It appears to be a failure to establish the objective degree of congruence between observed and expected findings (or elicited and extrapolated arrays).
18. The conditions of the error are not identified in the present study, but the nature of the diagnostic problem seems an ideal environment in which the error of set may occur.
19. Many features identified in the account gathering study are variations on the central characteristics of the second stage of the diagnostic thinking process, and do not themselves characterise that stage. Such variants are the specific errors of set, failure to make specific or general enquiry, failure to make the correct interpretation and inappropriate designation of irrelevance. Such errors arise from either inadequate knowledge or inadequate structuring of elicited or stored knowledge. Errors are not systematic, but arise unpredictably. It may be more productive, therefore, to identify conditions which preclude rather than encourage them.
20. The second characterising feature of the second stage of the diagnostic thinking process is the assessment of interpretations by working from the extrapolated context. This is analogous to the stage of hypothesis testing of the Michigan and McMaster theories.
21. The need to assess interpretations causes the problem solver actively to determine the course of the interview. Information collection is based on the expectations which arise out of the extrapolated contexts. This is the determinative role.

22. The determinative role is tempered by the reciprocal activity of responding to the sometimes unpredictable and partially uncontrollable information elicited as a result of interpretative and re-interpretative operations.
23. A certain threshold of data acquisition is a necessary but not sufficient condition of the diagnostic thinking process. Homogeneity of registrars' skills structures substantiates this inference.
24. The second stage of progress of the diagnostic thinking process is characterised by restructuring and by assessment of interpretations by working from the extrapolated context. The problem solver displays reciprocal determinative and responsive cognitive operations which reflect assessment and restructuring, respectively. Other cognitive operations of the second stage are variants on these central characteristics.

12.2.3 Stage Three: Resolution

The third stage of the diagnostic thinking process concerns the final selection, rejection or not of the interpretations and re-interpretations of the first two stages. It does not concern an unalterable endpoint, but only the point of greatest specificity and certainty which the thinker is willing to reach on the basis of the available, elicited information. At this stage, therefore, the thinker makes his judgment. Hoffman (1960) expresses this quite clearly:

"The primary task of clinical diagnosis is that of collecting, evaluating and assimilating information with respect to the patient. The starting point is the information itself ... The end result is a judgment".

The need for judgment rather than objective manipulation of indubitable data arises from the nature of the task and the problem. Armstrong (1977) considers that:

"The task of diagnosis is essentially the approximation of the patient's pathology to an established disease category. This process relies on sampling the manifestations of pathology

through examination and laboratory investigations and also through the patient's experience. But the patient's experience will rarely correspond directly with the formal indicators of internal pathology; it must be interpreted".

Cohen (1972) quotes Osler's dictum that "Medicine is a science of uncertainty and an art of probability" and specifies two categories of uncertainty which necessitate judgment rather than allowing objective data manipulation. These two categories of uncertainty Cohen (1972) calls intrinsic and extrinsic:

"Intrinsic uncertainties are those pertaining to the imprecision, ambiguity or other limitation of the data on which a medical judgment is to be based ... Extrinsic uncertainties, often misleadingly called "observer error", make their appearance in the interpretations given to a set of data. Here the interpretation may vary from one assessor to another, while one and the same assessor may, on a second occasion, give an interpretation which differs from the one he gave on the first". (p. 93)

Extrinsic uncertainty is, therefore, our main concern when considering the nature of the diagnostic thinking process from a psychological point of view. Selection among interpretations and its judgmental nature may be seen in the light of Peel's (1971) work. He considers that situations which call for "the intellectual resolution of possibilities and actualities" and which are "open in that several explanations may be possible" require the thinker to select by a process of judgment. He clearly defines and elaborates on the meaning of judgment in the following way:

"Judging is a form of thinking and is therefore invoked whenever we are in a situation for which we have no readymade answer learned off pat. But in addition, judgment refers to a situation for which there is no single final correct response to be discovered, but rather a spectrum of responses satisfying different numbers of different criteria. Some decisions, therefore, may be better than others - on certain grounds - but in extreme conditions none is outstanding and the decision turns on what the judger wants to fulfil". (p. 19)

The psychological nature of this phenomenon in the third stage of the diagnostic thinking process is the subject matter of this section.

The account gathering study provides data and discussion relevant to the third stage. Section 11.6 notes that selection among competing interpretations of clinical information is made either by active confirmation, or active elimination, or passive elimination (although this cannot be studied by the account gathering method), or selection is postponed because of insufficient or unreliable information. However, although each of these choices may represent the resolution of the thinking process, none is irreversible. Each, therefore, may be a resolution or a point of progress. The changing array of information may determine the nature of resolution as much as the thinker's own cognitive processes or style. In addition, the logical characteristics of the diagnostic problem ensure that a decision made about one interpretation will have consequences for the fate of other interpretations of the same or an overlapping array of information. This is discussed fully in section 11.6. It is also mentioned that, like Elstein et al (1978), we consider that clinical data are evaluated in terms of their fit to anticipated findings. The nature and origin of these anticipations and the mechanism whereby information is elicited are discussed in the previous two sections. However, to describe the process merely as one of fitting or comparing observed and expected findings seems to reduce it to a somewhat oversimplified form. This is particularly so given our preliminary discussion of the meaning and necessity of judgment in the diagnostic thinking process and the differences in criteria that may exist between judges.

De Groot (1965) terms the decision to reject a proposition the exhaustion point. His study of chess playing enables him to state that "rejection is a last resort which is only called upon when the means of strengthening run dry", although he also mentions the process of "negative proof" or deliberate hypothesis refutation. Where the exhaustion point is reached is, for de Groot's subjects, dependent upon the phase of the thought process. The earlier the phase, the sooner the

exhaustion point is reached. However, our data and the nature of the diagnostic problem indicate that such specificity of description is inappropriate. The descriptive framework of the phase succession has already been rejected (section 12.2.2) and the multiplicity of interpretations makes unlikely the evaluation of interpretations singly. Such a process as this may occur in a particular individual on a particular occasion, or for a particular hypothesis, but it is only one of many possible tactics or cognitive operations, as our data show. Other workers define other operations and approaches which may be, on any one occasion, of equal validity or descriptive accuracy for the diagnostic thinking process.

Bruner et al (1956) speak of the "inability or unwillingness" of their subjects to use negative information and suggest that this must usually be transformed into positive information resulting in risk of error. Although Donaldson (1959) does not accept this explanation in all circumstances, she also notes the phenomenon. Bruner et al (1956) describe four ideal strategies in concept attainment. The simultaneous scanning strategy assimilates all information and successively rejects impossible hypotheses. Successive scanning tests only one hypothesis at a time. Conservative focusing and focus gambling concern attribute changing which is not appropriate to the structure of the diagnostic problem. Again, either simultaneous or successive scanning may be descriptive of the diagnostic thinking process of one individual at one time. But neither necessarily describes the overall tendency of that individual. As is noted earlier, the logical characteristics of the diagnostic problem and the necessary inter-relatedness of hypotheses makes some forms of interpretation evaluation objectively impossible.

As with the discussion of the second stage of the diagnostic thinking process, we find the psychological literature helpful in identifying variants of the process, but not its central characteristics. We have so far identified only that observed and expected findings must be compared. Interestingly enough,

de Groot (1965) rejects such a formulation:

"A psychologically interesting question is whether we must think of this process as one of 'comparing' the two values, expectancy and outcome. Such a description does not appear adequate though: it is too intellectualistic. We can rather say that the subject is analysing variations with an anticipated degree or level of satisfaction in his mind".
(p. 259)

This subjective feature is discussed above and examples given of subjects' feelings of satisfaction (see section 11.6). Yet such satisfaction is often expressed precisely in terms of judgment of the 'fit' of observed and expected findings. Such judgment might be explained in terms of the central concepts of structure and extrapolation, yet these seem tied particularly to the first two stages. The third stage concerns the thinker's activity after structuring, restructuring and extrapolation are complete, if only ephemerally so. Failure to fit implies incongruence of the observed and extrapolated structures. This may arise, for example, if information elicited subsequent to the initial identification of the forceful feature or cue throws doubt upon that identification, but is not sufficient to cause restructuring. Thus observed and extrapolated structures no longer are potentially the same, but no new extrapolated context is yet identified. One could invent by speculation and theory many such circumstances which may cause the subjective impression of lack of fit. However, the only generalisable statement which may be made is that feelings of dissatisfaction which are not sufficiently defined to allow rejection of an extrapolated context imply an inability to restructure, for whatever reason, while the increasing array fails to replicate the extrapolated context.

A major problem for the third stage of the diagnostic thinking process is to define how much of the expected information must be observed before an extrapolated context is accepted as the true explanation of the observed information. A second problem is to define how, or according to what criteria, the thinker selects among extrapolated contexts which intersect.

Elstein et al (1978) have defined a series of judgmental rules (see 3.1.1) but these are not generalisable. Examples of such rules are: "At least three out of the following five features must be present to make a diagnosis of Y" or "X is diagnosed if and only if, all of the following features are present". These may be applicable in certain instances but, unless every extrapolated context is accompanied by such a rule, they can be only a variant on a central characteristic. The present study provides no specific data to resolve these possible features of the diagnostic thinking process. In this, we echo Elstein et al (1978) that:

"Unfortunately, this seems to be about as far as cognitive psychology can take us at the moment. A set of generalisations that indicates when each rule is, or should be, employed is not yet formulated".

The broadest generalisation we may make is that subjects confirm, eliminate or postpone judgment on their competing interpretations of the clinical information, and they do so with greater or lesser degrees of subjective certainty and satisfaction. We may add that subjects seek certain features and items of information not in isolation, but in relation to others. Thus, the reputed configurational ability of the clinician (see section 2.1) is reflected and substantiated in the present structuralist approach.

Finally, we must return to question of judgment. Section 11.11 discusses the implications of the account gathering study for our understanding of the nature of psychological probability. The data do not permit any more specific conclusion than that subjects tend to weigh up and balance data, assigning relative importance to contradictory items and reaching a conclusion of some degree of definiteness. The indirect and mediated nature of clinical evidence, the imperceptibility of symptoms and the vagaries of language ensure that absolute certainty of a complete clinical picture is a rare phenomenon.

The argument of this section may be summarised as follows:

1. Judgment of congruence of observed and expected information even though the observed array is less complete than the expected (or extrapolated) one is apparent.
2. Such judgment is amenable to reversal.
3. Resolution occurs by active confirmation, active or passive elimination or postponement of judgment.
4. Characteristics of the changing array of information and of the thinker's cognitive processes and structure determine the nature of resolution.
5. Resolution occurs within an extrapolated, structural context.

Psychological theory is not found very helpful in illuminating further the nature of these characteristics.

12.2.4 The Central Characteristics of the Three Stages of the Diagnostic Thinking Process

Crutchfield (1972) summarises in clear terms the complexity of the problem solving process, and the dangers inherent in describing it:

"... the problem solving process is not a cut and dried sequence of steps executed in the orderly manner of the pre-established programme of a computer ... the process of solving problems is highly complicated, involving a great number of separate but inter-related cognitive skills. Whether or not the individual succeeds in solving the problem depends heavily upon his ability to marshal these skills in an effective co-ordinated way. He must know which step to take at a given point, which plan of attack to adopt. He must know how to employ sensible 'stop strategies', for instance, to be able to decide when it is time to stop collecting information and to start generating ideas, when it is time to stop generating ideas and to start evaluating them, and when it is time to abandon one direction of thought and to embark on another. He must be able to balance and harmonise the often contradictory demands which arise in the problem".
(p. 195)

Although these processes may not be as discrete or conscious as Crutchfield seems to imply, his statement reflects that

made at the beginning of ^{section 12.2}~~this chapter~~. The differentiation of stages is merely a device for description and understanding. In reality an individual may be at all three stages simultaneously in the different threads of his thinking process. The boundaries between the stages are blurred and the same cognitive operation may be identified as belonging to more than one stage, depending upon its meaning for different interpretations of the clinical data.

Bruner (1951) also addresses the problem of describing psychological processes. His context is perception, but his argument is relevant to the present discussion:

"The reader may object that our model of the information confirming cycle seems too saccadic, too jumpy, that perception seems to work more smoothly than our model indicates. There are two legitimate answers to this objection. The first is that only under well practised conditions of perceiving is the process so smooth ... But this rejoinder is trivial in the light of the second one. There need be no phenomenal resemblance, we would insist, between the feeling tone of a psychic process and the conceptual model used to predict or describe it. Nobody would seriously object today, for example, that the atomic theory of matter is an inadequate theory because matter, a rock for instance, does not look or feel like an amalgam of whirling atoms".

Such is our evaluation of and approach towards the present theory and description of the diagnostic thinking process. But such qualifications must not be taken to imply that the model is paramorphic. We contend that it is isomorphic and descriptive and that the features and characteristics which are identified are real. The qualification refers only to the necessary formalised presentation of those characteristics which are suggested as follows:

Stage One: The Initiation of Interpretations is a sine qua non of the diagnostic thinking process, mediating between information elicited and diagnosis made.

Interpretation of information is achieved by extrapolation from the given information to arrays of knowledge stored in the

thinker's memory. This enables the thinker cognitively to structure (or arrange) the given items of information in one or many meaningful, and possibly correct, ways. Some form of cognitive selecting operation is hypothesised which enables the thinker to identify the piece or pieces of information in the elicited array which have the greatest payoff in terms of facilitating interpretation by extrapolation to useful stored embedding contexts. The operation is suggested as an available rather than necessary stage. The piece or pieces of information identified by means of the selecting operation are called the forceful feature. Different pieces of information have different interpretative value or forcefulness in terms of facilitating extrapolation to useful embedding contexts. The selecting operation may yield identification of a forceful feature which is not actually the cue (or true forceful feature) of the array, in which extrapolation to an inappropriate embedding context may occur since elicited and stored arrays of information, it is suggested, are most efficiently and effectively related in terms of their respective forceful features. Despite this, it is possible to extrapolate from any given piece of clinical information, whether or not it is identified as the forceful feature, but only with a greater or lesser degree of efficiency, effectiveness, facility and personal satisfaction. Thus, items of information must attain a certain threshold of force before being perceived as useful in the efficient or effective diagnostic thinking process. Threshold of force may thus be an absolute, not relative, value. The selecting operation facilitates active identification of the forceful feature by selective attention rather than passive acceptance of force according to the features of the presentation by field dependence. More than one forceful feature may be identified in any array of information and the same information may be the forceful feature embedded in many extrapolated contexts of stored, structured information. The major characteristics of the first stage of the diagnostic thinking process, then, may be summarised as:

Different degrees of forcefulness of items in the array.

Application of a selecting operation to identify the forceful feature of the array.

Interpretation by extrapolation.

Structuring elicited information.

Relating elicited and stored arrays by means of respective forceful features.

Possible extrapolation from any item in the array.

Degrees of efficiency, effectiveness, facility and personal satisfaction of extrapolated contexts.

Absolute threshold of force.

Identification of multiple forceful features in one array.

Multiple extrapolations from the same forceful feature (also characteristic of Stage Two when occurring over extended time).

Stage Two: Progress of the Diagnostic Thinking Process concerns the cognitive operations applied to the interpretations of Stage One. The stage is characterised by restructuring and re-interpretation of the same or the changing array of information from which the initial, Stage One, interpretations were made. Progressive specification is not a necessary characteristic of restructuring. Restructuring requires perception of separate, constituent elements of the information array as a necessary, but not sufficient, condition of the diagnostic thinking process. Restructuring can arise by means of shift of force or by extrapolation from the forceful feature to multiple embedding contexts. The occasional tendency has been noted to rationalise, ignore or find explanation for incongruent information in order to maintain a favoured extrapolated context. This is termed the set effect. Further error might arise if an inappropriate set is extrapolated from either the cue or an inappropriate forceful feature. The possible appropriateness of interpre-

tations is assessed by the thinker working from the extrapolated context and playing a determinative role in eliciting information. This is tempered by the reciprocal activity of responding to the sometimes unpredictable and partially uncontrollable information which is actually elicited. A certain threshold of clinical data acquisition is a necessary but not sufficient condition of the diagnostic thinking process. The major characteristics of the second stage of the diagnostic thinking process, then, may be summarised as:

Restructuring and re-interpretation of the changing or unchanging array of clinical information.

Restructuring by shift of force or by multiple extrapolation from the same forceful feature (also characteristic of Stage One when occurring in close contiguity of time).

Assessment of interpretations by working from the extrapolated context as the method of establishing congruence of elicited and extrapolated assays.

Acquisition of a sufficient threshold of clinical information elicited.

Reciprocal determinative and responsive cognitive operations and behaviour in relation to the flow of information.

Idiosyncratic but not systematic errors.

Occasional tendency towards the set effect causing error if an inappropriate extrapolated context is located from the cue or from an inappropriate forceful feature, resulting in failure of the process of establishing congruence between elicited and extrapolated arrays.

Stage Three: Resolution concerns the final selection, rejection or not of the interpretations and re-interpretations of the first two stages. It does not concern an unalterable endpoint, but the point of greatest specificity currently possible for

the problem solver. The major characteristics of this stage may be summarised as:

Judgment of congruence of observed (elicited) and expected (extrapolated) arrays of information, although the observed array may be incomplete or uncertain.

Possible reversibility or refinement of the judgment.

Resolution by active confirmation, active or passive elimination or postponement of judgment of interpretations.

Nature of resolution determined by features of the information elicited and the thinker's cognitive processes and structures.

Having thus traced the changing characteristics of the fundamental diagnostic thinking process through its three stages, we may now turn to variants on that fundamental process by considering its developmental and comparative aspects.

12.3 Developmental and Comparative Aspects of the Diagnostic Thinking Process

This section draws on the comparison of results of the questionnaire studies in endocrinology and neurology, and of the relative performances of the groups of subjects in each study. It is argued in Chapter Five (section 5.6.1) that the subjects responding to each questionnaire may be taken as a longitudinal sample. Section 5.6.2 argues the same case for subjects of the account gathering study. Section 5.6.3 establishes that the subjects of the two parallel studies are also comparable. We may, therefore, infer a description of the development of the diagnostic thinking process from the final year of undergraduate medical education through to the years at registrar level of clinical practice. We may also draw reasonable conclusions concerning the generalisability of that process across different clinical specialities.

In section 8.2 are suggested some differences between students' and registrars' experiences which may influence the nature

and therefore the development of the diagnostic thinking process. . These differences are cited as amount and type of exposure to clinical phenomena (interviewing and examining patients and making diagnoses primarily in order to learn, as opposed to undertaking the same processes primarily to manage, treat and cure the patient as part of the workload of clinical responsibility); contextual aspects of knowledge and skill development (learning as a student as opposed to learning as a clinical practitioner); recency of initial knowledge and skill acquisition; relative rehearsal of knowledge and skills; contextual aspects of knowledge and skill rehearsal (rehearsal in order to become proficient, rehearsal in order to maintain proficiency, rehearsal as a response to a particular clinical situation, etc.); relative decay of unrehearsed knowledge and skills; and changes in the storage structure, inter-relatedness, use value, meaning or significance and corollaries of knowledge and skills resulting from practice. From subsequent discussion, we may now infer that such differences as these may affect the structure of stored information, the selection of forceful features, facility of extrapolation, flexibility of interpretation and the cognitive skills of the second and third stages of the diagnostic thinking process.

The questionnaire study, by indicating differences between students and registrars, also allows inferences about a developmental process. In endocrinology, it is shown that students and registrars have relatively similar skills structures but different cognitive processes and, not surprisingly, different diagnostic acumen (see Table 8.1). In making a diagnosis, students rely much more heavily than do registrars on the skills measured by the predictor sections of the questionnaires (41 per cent and four per cent of the variance, respectively). These skills are mastery of factual knowledge, interpretation of symptoms and signs and selection of diagnostic possibilities, tests and investigations. It is therefore concluded that the cognitive process skill develops with clinical practice. However, the students display skills

structures similar to those of the registrars, but, clearly, at a lower level of achievement. Viewing these findings developmentally, we may say that in endocrinology the neophyte differentially develops the substance of that speciality and later develops the aptitude.⁽¹⁾ In other words, process is built on content. The two aspects of cognition do not appear to develop simultaneously in their structure. The elements precede the processes (see 12.1 above).

Turning to neurology, the questionnaire study shows that students and registrars have similar cognitive processes, but different skills structures (see section 8.4.3) and, predictably, different diagnostic acumen. In making a diagnosis, students rely to the same extent as registrars on the skills measured by the predictor sections of the questionnaires (27 per cent of the variance accounted for in each case). Thus, the cognitive process skill in neurology, unlike in endocrinology, in undergraduate medical education already resembles that of clinical practice. However, the content associated with those processes does not develop the relative characteristics of clinical practice until after the years of undergraduate medical education. Thus, in developmental terms, unlike endocrinology, the student differentially develops the aptitude of the speciality of neurology prior to developing its substantial structure. As with endocrinology, however, the two aspects of cognition do not appear to develop simultaneously in their structures.

These findings suggest that development of the diagnostic thinking process in both specialities is characterised by increasing equilibration of skills structures, possibly beyond a certain threshold level (see section 8.8). This is so even for endocrinology, in which speciality a non-significant interaction term (section 8.4.3) suggests similarity of students' and registrars' skills structures. But this similarity is relative and not absolute (see section 8.6.1), we may only say with justification that endocrinology yields a relatively less differentiated set of scores than does neurology (section 8.6.3).

Footnote (1) - see end of chapter

However, it is worth noting that this is only true when scores on section D (formulating a diagnosis) are included. If these are excluded, the results of the Scheffé tests (Tables 7.19, 7.21, 7.23 and 7.25) show that in endocrinology students and registrars are relatively less similar than they are in neurology. Thus the importance of the role of the untested skills is indicated. In endocrinology, there is an increasing development of, reliance upon or use of these skills with experience of clinical practice. This does not appear to be so in neurology. Many of these developmental differences are attributed to the different effect of medical education in the two specialities (section 8.8), but it remains the case that tendency to equilibrate the skills structure is a fundamental characteristic of the development of diagnostic skill and that the development of a relative balance of contribution from those skills and others not tested appears to be speciality specific. Thus, it is shown that cognitive processes but not skills structures are speciality specific (section 8.9).

However, such a statement requires clarification, since it is based entirely upon the predictive value of the first three sections of the questionnaire for scores on the fourth and on the amount of variance left unaccounted for in this prediction. In stating that cognitive processes are different we are really only saying that different amounts of variance remain unaccounted for. Thus, they are different for students and registrars in endocrinology and for registrars in endocrinology and neurology, while being less dissimilar for students in endocrinology and neurology and similar for students and registrars in neurology. But are we to infer from this that such similarities are qualitative or quantitative? The account gathering study provides evidence in relation to this question.

It will be recalled (from Chapter Eleven) that in all the categories identified, no statistically significant differences were found between the groups of students, house officers and registrars. We must conclude from this that students are in

command of a range of cognitive skills not different from that of registrars. Whether the presence of each of these skills is a function of medical education or cognitive development generally, is a matter for speculation, but it would seem reasonable to infer that aptitude itself (i.e. cognitive processes such as structuring, extrapolation, tendencies to confirm or reject interpretations, etc.) is a function of broader cognitive development and that aptitude is identified, here, expressing itself or operating with a certain substance. Thus, for example, the tendency to structure the world is a general characteristic of human cognition (see 12.1.1), however, the development of particular structures which may be located with greater or lesser facility is a function of experience, in this case of medical education and clinical practice. In essence, it appears that students, taken as a group, display all the cognitive processes of house officers and registrars. If this is so, wherein lies the difference between tyro and expert?

One difference certainly lies in breadth, depth and extent of knowledge. More importantly, it lies in the storage structure and use value of that knowledge. Section 4.1.3 points out that problem solving itself is a form of learning and, therefore, a spur to cognitive development in that working with information and applying it will have modifying effects upon the storage structures of that information in memory, upon its accessibility and manipulability. In the present case, the selecting operation may become more efficient and effective with experience but that does not suggest that a less efficient and effective operation is not applied to the very first clinical problem encountered. Likewise, information learned in lectures or textbooks or laboratory may assume a different meaning and significance or may lose meaning and significance, when considered in the light of clinical practice. By this, we are saying no more than that its structural characteristics alter. As structure alters, so must the substance and the efficiency and effectiveness of extrapolation.

.

The account gathering study shows that students are less well equipped than are house officers and registrars to "think on their feet" (section 11.8). It is suggested that this cannot be a failure of knowledge as such (since the failure is identified by the subject himself) but is a failure of structure or of extrapolation due, perhaps, to the way in which students have learned and stored information and its relationship to the structure of information elicited during the clinical interview and the ability to impose structure on that information and locate related information in the stored structure which must be sought in the problem field. Thus, the developmental characteristics of the knowledge base concern not only the substance itself or its quantity but also its structural properties and the changes in these with experience and use. The developing ability to "think on one's feet" may reflect an increasing definition and stability of the extrapolated structure.

With regard to aptitude, it is reasonable to infer that, account gathering results notwithstanding, there are developmental differences. The results of the questionnaire study indicate that, for endocrinology, there is an increasing, developing reliance upon the processes not tested. Although the account gathering study shows the same cognitive operations and processes in students, house officers and registrars, it does not demonstrate their relative distribution within each of these groups. These two findings may suggest that the development of such cognitive operations in the diagnostic thinking process is, therefore, not a matter primarily of qualitative change, but rather of quantitative change. Thus, for example, the neophyte may emphasise extrapolated interpretation of all items of information, while the expert may prefer to exclude some items of information as irrelevant. There may be differences in terms of willingness to re-interpret information or to reject interpretations without having performed an exhaustive search. These are merely examples and do not purport to be real, but the differing relative prominence or incidence of the cognitive operations identified is put forward

as a developmental characteristic of the diagnostic thinking process. Just as it is suggested that the structure of stored knowledge changes with experience, so is it also now suggested that the structure of aptitude also develops and changes. This possibility is mentioned in 12.1.1 above. It is emphasised, however, that the processes themselves do not change.

Although a general developmental model of the diagnostic thinking process is thus postulated, it is necessary to consider whether or not speciality specificity, as shown in the questionnaire study, restricts its applicability. It is reasonable to suggest that it does not. Given the generalisability of the elements of the model to other areas of problem solving and learning (see sections 4.1 and 12.1) and their roots in general psychological theory, it is tenable that general cognitive processes have been identified which apply equally to other substance. Thus they would apply in all specialities. However, the differences demonstrated between registrars in endocrinology and neurology make it reasonable to infer quantitative differences in cognitive processes across specialities, just as we have inferred this as a developmental feature. Thus different substance may be more efficiently and effectively dealt with by different cognitive operations. In terms of quality, however, we would suggest no differences between specialities.

The work of Ekwo and Lo ning-Baucke (1979) gives support to our contention that both substance and aptitude develop but that aptitude develops structurally or quantitatively rather than qualitatively. Ekwo and Loening-Baucke (1979) show that students in their study obtained 90 per cent of all data for both familiar and unfamiliar problems but that they did not always recognise its significance or use it to the full advantage. Thus both substance and aptitude were inadequate. Elstein et al (1978) also recognise this dual aspect and hint at the problem solving experience as learning by altering the structural aspects of memory. Their discussion, however,

is with reference to good and poorer problem solvers:

"The differences between experts and weaker problem solvers are more to be found in the repertory of their experiences, organised in long term memory, than in differences in the planning and problem solving heuristics employed ... This does not imply that medical problem solving is dependent solely upon mastery of passively recalled content. Knowledge must be retrieved and organised.
(pp 276-277)

Thus, the Michigan studies indicate a range of similar cognitive operations in expert and weaker problem solver, but point to the structure of substance as an important feature. Our own interpretation of the diagnostic thinking process, which gives pride of place to structural features of cognition, may cause us to reverse the order of Elstein et al's last two participles. Ianssek and Balla (1979) seem to be suggesting some differences between novice and expert which may be seen in terms of structure or interpretative activities as we have presented them. In discussing the proportion of fruitful (clarifying or hypothesis testing) questions asked by novice and expert they state:

"It must be pointed out that this parameter in no way took into account the capability of a novice's clarifying question to convert the symptom into a meaningful entity representing a neurological malfunction. In this context, the expert asked clarifying questions which usually achieved symptom conversion, whereas the student asked clarifying questions which merely elaborated on the surrounding circumstances of the symptom and failed to achieve conversion".

Although Ianssek and Balla (1979) do not describe the precise nature of symptom conversion, the context implies an operation similar to that of structuring and interpretation as used in the present study. In his study of chess players, de Groot (1965) noted the same developmental phenomenon:

"The rapid insight of the chessmaster into the possibilities of a newly shown position, his immediate 'seeing' of structural and dynamic essentials, of possible combinatorial gimmicks, and so forth, are only understandable if we realise that as a result of his experience he

quite literally 'sees' the position in a totally different (and much more adequate) way than the weaker player ... It is above all the treasury of ready 'experience' which puts the master that much ahead of the others. His extremely extensive, widely branched and highly organised system of knowledge and experience enables him, first, to recognise immediately a chess position as one belonging to an unwritten category with corresponding board means to be applied, and, second, to 'see' immediately and in a highly adequate way its specific, individual features against the background of the category".

'Seeing' in de Groot's terms, is analogous to the 'structuring' of the current study. The perceptual and physical circumstances of the game of chess make the description more immediate and comprehensible, but it is suggested that similar operations apply where the field is composed of verbal, tactile, quantitative, perceptual or descriptive information. De Groot's (1965) study of chess players also gives support to the theory of development suggested in relation to the present study:

"... a master is a master primarily by virtue of what he has been able to build up by experience; and this is: (a) a schooled and highly specific way of perceiving, and (b) a system of reproductively available methods in memory".

We may relate (a) to substance as structured in memory which yields in propensity, particularly by means of the selecting operation, to perceive and interpret the field in certain ways. We may relate (b) to aptitude or processes, although the present study would question the reproductive use of those processes. In general, however, de Groot (1965) makes relevant comment on the nature and consequences of the rather amorphous entity termed 'experience'. The close identity of problem solving and learning, the dual nature of the diagnostic thinking process and its structural development are thus reflected in the work of de Groot despite the very different nature of the subject areas of chess and clinical diagnosis.

The developmental and comparative aspects of the diagnostic thinking process thus seem to have a number of essential

characteristics which may be seen as arising from the structural aspects of the cognitive process. These characteristics may be summarised as follows:

1. The development of the diagnostic thinking process concerns changes in the structure of stored information, the selection of forceful features (i.e. alteration in the structure of the selecting operation), facility of extrapolation, flexibility of interpretation and the cognitive skills of the second and third stages of the process (see section 12.2.4).
2. The two cognitive aspects of the diagnostic thinking process (substance and aptitude) develop differentially in both endocrinology and neurology. In endocrinology, substance has preferential development, while in neurology, aptitude appears to have developmental predominance.
3. In both endocrinology and neurology, development is characterised by increasing equilibration of the skills structures as measured by sections A, B, C and D of the questionnaires.
4. The development of a relative balance of contribution towards the diagnostic thinking process from the skills tested in sections A, B and C of the questionnaires and others not so tested is speciality specific in endocrinology and neurology.
5. Speciality specificity appears to apply to cognitive processes, but not skills structures.
6. In the account gathering study, no statistically significant differences are found in any category between groups of students, house officers and registrars. It is reasonable to infer that all three groups appear to be in command of the same range of cognitive skills.
7. The cognitive processes identified are characteristic of broader cognitive development and are not particular to

the clinical problem solving environment in which they are identified. A general tendency of process is thus displayed in a particular context of substance.

8. The substance of the diagnostic thinking process develops in breadth, depth and extent, and in its storage structure and use value.
9. As the structure of substance alters, so does the concomitant efficiency and effectiveness of extrapolation.
10. Development of the diagnostic thinking process brings with it an increasing ability to "think on one's feet". It is suggested that this reflects an increasing facility in working in and with the extrapolated context or an increasing definition and stability of those contexts.
11. It is suggested that the development of aptitude in the diagnostic thinking process is not primarily a matter of qualitative change but is rather a change in quantity and emphasis. Different cognitive processes may predominate at different stages of development and experience but those processes themselves do not alter.
12. Points (8), (9), (10) and (11) suggest that development of the diagnostic thinking process is characterised by changes in the structure of both substance and aptitude.
13. The general developmental model postulated is not speciality specific, although quantitative (i.e. structural) differences in aptitude appear to occur between specialities. It is suggested that different substance may be more efficiently and effectively dealt with by different cognitive operations. However, qualitative aspects of the model are generalisable.
14. It is suggested that development is amenable to shaping by education.
15. The postulated developmental model of the diagnostic thinking process is substantiated in the work of Ekwo and

Loening-Baucke (1979), Elstein et al (1978), Iansek and Balla (1979) and de Groot (1965).

12.4 Summary and Conclusions

The results of the questionnaire and account gathering studies are related and a unified explanation and description of the diagnostic thinking process in undergraduate medical education and in clinical practice presented. The fundamental psychological features of that process are identified as structure and extrapolation. These are discussed separately in relation to the results of the parallel studies and to wider psychological theory. These discussions are summarised at the conclusion of sections 12.1.1 and 12.1.2 respectively. Three stages in the diagnostic thinking process are identified and discussed. These are: firstly, the initiation of interpretations; secondly, progress of the diagnostic thinking process; and thirdly, resolution. The discussion of each of these stages is summarised at the conclusion of sections 12.2.1, 12.2.2 and 12.2.3 respectively. A discussion of the nature of the model of the diagnostic thinking process thus postulated and a resumé of the central characteristics of three stages is presented in section 12.2.4. Section 12.3 concerns developmental and comparative aspects of the diagnostic thinking process. The discussion is summarised at the conclusion of that section.

Footnote on page 340

- (1) For the purpose of this discussion, the terms 'elements' and 'substance' are used interchangeably, meaning the content of thinking.

Footnote on page 394

- (1) For the purpose of this discussion, the terms 'aptitude', 'process' and 'cognitive processes' are used interchangeably and take Peel's (1967) usage, meaning 'thinking method' or 'problem solving processes' (see 4.1.2 above). In particular, no other psychological meaning of 'aptitude' (e.g. 'propensity', 'capability', 'ability') is implied.

CHAPTER THIRTEEN

Implications of the Current Study for Pedagogical Approaches in the Development of the Diagnostic Thinking Process

On the basis of the findings of the two parallel studies, the theoretical interpretation of these and the review of literature and pedagogical practice, we may now consider the implications of these for medical education. This chapter comprises three main sections. Firstly, we consider both the apparent and potential relationships between medical education and clinical practice with regard to development of the diagnostic thinking process. Secondly, aims for the design of teaching strategies are considered. Finally, a discussion of possible pedagogical approaches is presented and evaluated. Before proceeding, however, it must be stated that, having made no direct study of the learning and clinical experiences of our subjects, this chapter is based on second order inference and psychological and pedagogical theory. Its suggestions and conclusions, therefore, can only be tentative and propositional.

13.1 Medical Education and Clinical Practice

The relationship between medical education and clinical practice in the development of the diagnostic thinking process may be considered from two viewpoints: the relationship as it appears to be at present; and, the relationship as it potentially might be.

13.1.1 The Current Relationship

The current relationship between undergraduate medical education and clinical practice may best be assessed by its effects. Perusal of the first volume of the General Medical Council Survey (GMC, 1977) indicates a large number of small differences between the curriculum arrangements of the 38 medical schools of the British Isles and Eire. Large differences in curriculum organisation and philosophy or in teaching methods are described and discussed in Chapter Four above. However, apart from the present study, no data or results have been available which facilitate some estimation of the relative effects of

medical education and clinical practice on the diagnostic thinking process.

The current questionnaire study yields a number of findings which are of apparent importance to a discussion of the relationship between medical education and clinical practice. In effect, the relationship appears not to be a complex one. Firstly, that undergraduate medical education appears differentially to enhance or facilitate the skills tested and those untested is made clear by the results of the one-way analyses of variance and Scheffé tests (see sections 8.3 and 8.6). This interpretation is justified in the light of the second finding of the equilibrating or compensatory effect of clinical practice on those skills, such that scores on the four sections of the questionnaire for registrars are not statistically significantly different (section 8.6). Thirdly, the shaping effect of undergraduate teaching and learning on thinking processes is also made clear by the results of the linear regression analyses. In endocrinology, students and registrars are very dissimilar; while in neurology they are very similar (section 8.3).

It is clear, then, that medical education has a definite effect on skills facilitated, and that this effect is different in the two specialities (see section 8.6.3). In addition, the effect produced by medical education is unlike that of clinical practice, either in skills structure or cognitive processes, in both specialities. The presentation of each speciality yields in students a cognitive structure (skills structure plus cognitive processes) particular to itself. Although that structure changes with clinical practice, each speciality remains distinct in terms of the cognitive structure of its practitioners in that registrars in endocrinology and neurology rely to different extents on the tested and untested skills when making a diagnosis. However, homogeneity of level of skills as measured by the four sections of the questionnaire, characterises registrars in both specialities, regardless of the skills structures shown at the end of undergraduate medical education (see section 8.8). A discussion of the possible derivation of those differences is presented in section 8.8.

in terms of the different approaches of the specialities studied and the consequent different teaching approaches. Endocrinology is described as having an undifferentiated, unanalytical or holistic approach, whereas neurology has a clearly articulated and logical philosophy of diagnosis and teaching approach. The former yields a non-statistically significant interaction term between scores of students and registrars; the latter yields similar reliance on the skills tested in making a diagnosis but less relative effectiveness in actually achieving that diagnosis correctly.

The speciality differences shown by the questionnaire study are particularly interesting when we recall that the student subjects would all have had experience in both endocrinology and neurology, and would have been taught according to their respective pedagogies. It would appear, therefore, that an approach learned in one speciality is not necessarily appropriate or transferable to another speciality. This is not to suggest, however, that cognitive operations as such are actually different. Indeed, we suggest that thinking processes are qualitatively similar although quantitatively different across specialities. To what, then, may we attribute the differences between students in the two specialities?

It is reasonable to assume that the difference is found in the dominant organising concepts of the two teaching approaches. The approach of neurology seems to be process orientated (see section 8.8) whereas the approach of endocrinology teaching, not having a clearly articulated logic of diagnosis, is by default content orientated. We, therefore, see that in each speciality the less emphasised aspect (process in endocrinology; content in neurology) is developmentally disadvantaged in the diagnostic thinking of the student subjects. It seems hardly surprising that students rely primarily on the skills which are emphasised to them. Thus, both endocrinology and neurology teaching achieve in students a relationship between skills structures and cognitive processes which is not reflected in clinical practice. However, the account gathering study shows

quite clearly that although a cognitive structure is being fostered in students which is not thoroughly appropriate in their later clinical practice, all the necessary cognitive skills are waiting in the wings.

To a certain extent Barrows and Bennett (1972) are correct in their assertion that diagnostic thinking skills develop only with clinical practice. Yet this needs qualification, for the present studies have shown that those skills are all present in the student, as much as the house officer or registrar. Further, each process skill may well be fully fledged, although it is clearly untenable to suggest that the substance requires no elaboration. What develops with clinical practice seems to be a particular inter-relationship or structure of these skills. It is this aspect which medical education leaves in embryonic form only. In section 12.1.1 it is argued that the stored structure of knowledge (both aptitude and substance) gained in medical education must, perforce, be unlike the stored structures which result from clinical practice, since the contextual aspects of the two are, and must be, different. In sections 11.8, 12.1.1 and 12.1.2 it is suggested that the effect of clinical practice is to alter the stored structure of aptitude and substance gradually rendering it more appropriate, complete and accessible in terms of clinical practice. This may also imply that it becomes less appropriate, complete and accessible in terms of the formal requirements of medical education. Sections 8.2 and 12.3 mention some conditions of clinical experience which may facilitate such structural alteration and development. It is self evident that such alteration is a function of learning and that clinical problem solving itself is its vehicle or instigator during the years of clinical practice. The earlier discussion of problem solving as learning gives support to this view (see section 4.1.3). Lovell (1961) suggests that experience is the greatest developer of thinking skills. This is echoed in the advice of Wyn Pugh et al (1975) that, since rapid improvement follows qualification to practise medicine, the factors in the postgraduate years which lead to such a change should be

identified and incorporated into the undergraduate curriculum. They presume that this principle "underlies the policy of those medical schools in which final year students take increasing clinical responsibility".

In conclusion of this section, we may summarise the current relationship with regard to development of the diagnostic thinking process between medical education and clinical practice as follows:

1. In the specialities of endocrinology and neurology, undergraduate medical education facilitates either a relatively appropriate skills structure (endocrinology) or relatively appropriate cognitive processes (neurology) in relation to the experienced practitioner. In neither case are both aspects developed to similar degrees of appropriateness.
2. The different teaching approaches in endocrinology and neurology foster different cognitive structures in students.
3. Structural development of the aptitude and substance of the diagnostic thinking process occurs with clinical practice due to the increment in learning which accrues from problem solving.
4. Homogeneity of skills is characteristic of the experienced practitioner in both endocrinology and neurology, but the relative contributions of these towards the diagnostic thinking process is different.
5. Students have all the cognitive skills of the experienced clinician, but use them in different proportions. Thus, with regard to aptitude, development of the diagnostic thinking process is primarily quantitative.
6. With regard to substance, clinical practice facilitates both qualitative and quantitative development of the stored structures.
7. As stored structures of aptitude and substance become more appropriate, complete and accessible in terms of clinical practice, they may become less so in terms of the formal requirements of medical education.

13.1.2 The Potential Relationship

In considering what the relationship between medical education and clinical practice should be, we must bear in mind three factors.

Firstly, we have shown that the diagnostic thinking process facilitated during undergraduate medical education is not commensurate with that demonstrated in clinical practice in that some aspects of that process develop preferentially. Secondly, it is also shown that the separate skills of the diagnostic thinking process are in evidence during the final phases of undergraduate medical education. We may suggest that this is due to two factors: firstly, necessity to teach and learn about the facts of bodily function and malfunction, the indicators of disease and health and the available range of diagnostic tests and investigations; and secondly, normal cognitive development which brings with it a certain range of possible cognitive operations which may be applied to the problem of making a diagnosis just as much as they may be applied to any other problem of similar complexity in any other sphere of life. Thirdly, it is shown that development of the diagnostic thinking process continues during the years of clinical practice addressed by this study.

Against the background of these three findings, it is necessary to consider some fundamental issues. The first of these concerns whether or not a student should be equipped with the same skills, in the same proportions as the expert practitioner, or whether it is reasonable and safe to allow some skills to develop or refine with practice and application after the stage of basic training (here, undergraduate medical education). If the latter option is selected, then what criteria are to be used to decide which skills will be allowed mainly to develop with practice? In addition, what criteria will be applied to determine the degree or level of basic skill necessary in order to benefit most advantageously from the experience of clinical practice?

The finding of an apparent compensatory effect of clinical practice on the skills less preferentially facilitated by undergraduate medical education allows us to consider the question in terms of other issues. For example, one of two viewpoints could be adopted. Firstly, we could conclude with some apparent justification that the ability to compensate for deficiencies is engendered, or at least not dispelled, by undergraduate medical education and therefore the status quo is acceptable and appropriate. Or, secondly, we could conclude that there should be no necessity for compensatory activity, and therefore the status quo is unacceptable and inappropriate. In essence, either of these positions is untenable for the same reason, that medical education can only be a preparation for practice and is, of its very nature, different from practice. Education should, at best, facilitate the basic skills and knowledge necessary for practice. This being so, it is a defensible proposition that, with regard to the variables here tested, medical education is quite appropriate, since the skills engendered in education are all used in practice varying only in their relative magnitude of use. What is perhaps not defensible, however, is that a vast area of relevant skill is not specifically addressed by the educative process. That area is reflected in the variance on section D of the questionnaires unaccounted for by sections A, B and C in all groups of subjects and consists of the cognitive processes identified in the account gathering study and summarised in relation to the three stages of the diagnostic thinking process (see section 12.2.4).

In response to these issues, it must be reiterated that medical education is a preparation for clinical practice and is not the same as practice itself. It therefore can produce a doctor who is qualified to practise but is not expert or experienced in doing so. This is widely recognised at a number of levels. For example, in terms of the rapid growth of medical technology, Parry reports in 1978 that:

"... it was no longer possible for undergraduate education to produce a doctor fully qualified to practise the whole range of medicine. All doctors needed postgraduate specialist education during the early years of practice, as well as continuing education during the rest of their professional lives". (Parry, 1978, p.3)

This opinion reflects that of Barrows (1973). It is clear, then, that undergraduate medical education can only be a necessary preparation, but not a sufficient prescription for clinical practice. But what form should this preparation take? There are some useful, authoritative and complementary answers to this question:

"The object of this basic medical education should be to provide a basis for future vocational training; it is not to train doctors to be biochemists, surgeons, general practitioners, or any other kind of specialist". (General Medical Council, 1967).

Smart (1978) echoes this recommendation:

"It was no longer possible for the undergraduate course to encompass the whole of medical knowledge and to produce an all-round doctor; in these circumstances the aim of such a course was to train the minds of students, with medicine as the medium of education. The course should cover the basic knowledge, skills and attitudes of the profession and provide a certain amount of broad experience ...". (p. 5)

Only Barrows (1973) takes the process of diagnosis as the lynchpin of this general, basic and preparatory education:

"We can't teach all the content that may be appropriate in every subject of medicine ... The most appropriate gift we can give our students is the skill for finding the appropriate content he needs whenever he needs it in the future. Therefore, instead of being concerned about what the student learns about cardio-respiratory anatomy or physiology we should be more concerned that he is able to recognise the existence of cardio-respiratory problems, define the problems, derive accurate data about the nature of the problem and by the application of good self-study skills acquire the knowledge and information he needs to understand and manage the problem". (p.27)

Such aims as those defined by the GMC, Smart and Barrows, remind us of Bruner's (1960) contention that:

"The first object of any act of learning, over and beyond the pleasure it may give, is that it should serve us in the future. Learning should not only take us somewhere; it should allow us later to go further more easily".

(p. 16)

How, then, may the sought for potential relationship between medical education and clinical practice be defined? We find that it must be defined in its current terms.

Undergraduate medical education must be a broad preparation for clinical practice and the continued, additional learning which accompanies it. This conclusion is both a logical and a pragmatic one. It is logical in the sense that clinical problem solving makes continued learning (of aptitude if not substance) unavoidable (see 4.1.3 above). It is pragmatic in that undergraduate medical education is limited in time, students are limited in their capacity to learn given such time limits, and the body of knowledge and skill which could be learned far outstrips that capacity. Commonsense would suggest that this is true of any undergraduate course, student and subject.

We have, then, confirmed the status quo and concluded that undergraduate medical education must be, even in an ideal form, preparatory. However, it is necessary to define this more closely since realisation of that preparation, its content and organisation may take many different forms. It is not the sphere of this study and discussion to presume any definition of what constitutes the best subject matter or content of that preparation, for we are only concerned with the thinking processes and the nature and role of knowledge and cognitive skill associated with it. We are not concerned, here, with what that knowledge might be. We are concerned, however, with the structure of both substance and aptitude as they are learned and as they develop. We are therefore concerned with the presentation of the preparatory material

and its effects upon the development of the diagnostic thinking process. We are concerned with preparation to think as well as to learn and to know. It is indicated in the review and discussions (Chapters Four, Eight, Eleven and Twelve) that this is a matter worthy of some further consideration. The following section, therefore, initiates identification of aims and criteria in the design of teaching strategies to facilitate development of the diagnostic thinking process.

In summary of the present section, however, our conclusions are as follows:

1. Undergraduate medical education can only be a preparation for clinical practice. This confirms the status quo.
2. As a preparation, it should facilitate the basic skills and knowledge necessary for practice.
3. The vast area of cognitive skill which comprises the diagnostic thinking process must form part of the basic preparation. This constitutes a preparation to think which complements the preparation to learn and to know. It is therefore necessary to consider the structure and presentation of undergraduate medical education.

13.2 Aims in the Design of Teaching Strategies

This section concerns a definition of the general intended outcome of any teaching strategy designed to facilitate development of the diagnostic thinking process. It also necessarily concerns a discussion of the types of characteristic which such a strategy should display.

The results of the parallel studies indicate the complexity of the diagnostic thinking process and its manifestations in different specialities. Each study identifies specific aspects of the process. The questionnaire study identifies three skills to which undergraduate medical education may be said to address itself (mastery of factual knowledge; interpretation of symptoms and signs; selecting and testing diagnostic possibilities). The account gathering study identifies many

more cognitive skills and sources of error. In section 12.2.4 ten, seven and four cognitive processes are identified in relation, respectively, to the first, second and third stages of the diagnostic thinking process. It is also emphasised in section 11.5(above) that the desirability of any particular cognitive process cannot be judged in absolute terms, but only in relation to specific circumstances or cases. Thus, prescription of specific diagnostic thinking processes is neither appropriate nor reasonable. The dynamic, unpredictable and uncontrollable aspects of the clinical interview ensure that only generalisable statements of a very broad type can sensibly be made. In effect, such statements are analogous in generality to the eschewed description of "hypothesis generation and testing". We may say that structuring, extrapolation and interpretation are cognitive processes common to the resolution of all clinical problems, but the current studies indicate that few more precise statements may be made. Indeed, many such statements could only be made if the variables which give rise to certain processes, and the parameters which define the field and flow of information, are identified and defined. It seems doubtful that such variables and parameters could be identified in a sample of clinical interviews in any way which would permit generalisation to other circumstances. Even if this were possible, it might scarcely be of practical use to the clinical problem solver since it would only add to his task yet another level of analysis.

Given the multiplicity and variety of variables, circumstances and processes, and the range of possible interactions of these, it is concluded, then, that specific sequences of diagnostic thinking processes cannot be prescribed and incorporated into undergraduate medical education in any detailed specific or prescriptive form. Individual differences of cognitive functioning alone would preclude such prescription. However, the range of cognitive processes is amenable to definition, as the current studies show. These processes are

varied from the generalisability of structure and extrapolation to, for example, the specificity of the set effect. It is therefore suggested that undergraduate medical education should facilitate an awareness of those processes at an appropriate level of specificity, generalisability and practicality. It might not be necessary, for example, for clinical practitioners or students to have an awareness of the theoretical aspects of structure and extrapolation. Rather, it is important that both students and practitioners use the thinking and enquiring processes described in those terms to their greatest effect and efficiency. This implies that teaching and learning strategies must be developed and designed on the basis of those theories.

The conclusion that self awareness should be facilitated and not prescriptions imposed, is presaged in section 4.1.3 earlier. It is worth reiterating the point made:

"... the diagnostic thinking process must rely heavily upon each individual's own cognitive style and propensity. Externally imposed forms, such as method of data collection or recording, may have no effect upon the problem solver's actual thinking processes. To specify the particular properties of the clinical problem solving process, therefore, does not necessarily imply specification of a single or particular form, but may mean only the identification of a range of possible processes".

Thus, as Chapter Four concludes, teaching strategies must present the diagnostic thinking process as an entity amenable to conscious awareness and monitoring but not necessarily to particular shaping other than correction should cognitive error occur. Such a conclusion is supported by Lovell (1974) who suggests that the influences at work in aiding development of general ways of knowing include, firstly, education and culture which subsumes types of concepts used and discussed and opportunities to consider possibilities; and, secondly, auto-regulation or the reflection of an individual on his own co-ordinating activities. Lovell (1974) adds that the development of such general ways of knowing determines ways

in which particular knowledge is assimilated.

Given the current structuralist interpretation of the diagnostic thinking process, this is surely an aspect of primary importance. The nature and development of the cognitive structure of elements and processes will, it is suggested, be the ultimate determinant of degree of efficiency and effectiveness of the diagnostic thinking process. It is argued earlier (section 12.3) that change in the structure of both substance and aptitude is a major developmental characteristic of the diagnostic thinking process.

Barrows (1973) seems also to have noted this phenomenon but appears to imply that development is merely addition of information and not change in its structure. Speaking of the developing clinical problem solver he writes:

"As his experiences enrichen, his hypotheses should become multiple memory associations which produce a greater range of possibilities for the patient's problem". (p. 22)

Barrows here reminds us, however, of the importance of substance. Eagerness to have clinical problem solvers develop and use an awareness of their own cognitive processes cannot be allowed to demote the necessary foundation of knowledge. Aptitude and substance must develop appropriately in tandem if the psychological interpretation of the diagnostic thinking process here presented is accepted.

Medical education has already been criticised for presenting factual information in a form unrelated to clinical practice. For example, Weed (1969) considers that if the teacher does other than:

"... present facts and principles as adjuncts of concrete experience, entering the student's mind at the level it understands ... he may erect a structure of abstract information in the student's mind that is confirmed by tradition, that passes for education, but that, when the student is confronted with a particular set of variables and needs to take meaningful action, may crumble and leave him anxious and confused".

A more frequently voiced criticism is that medical education primarily concentrates on the student's acquisition of factual information and an understanding of pathophysiologic mechanisms (Cutler, 1979), whereas sound clinical judgments are derived equally from these and the skill to combine them appropriately (Schwarz et al, 1973; Barrows, 1976; Myler-Crook, 1974). As Barrows and Mitchell (1975) point out, having a set of facts at his finger tips is no guarantee either that the student will know when they are important in his work, or that he will know how to use them. Indeed, Barrows and Bennett (1972) suggest that not only does traditional medical education not teach clinical problem solving skills, but actually inhibits its development by forcing students to learn systematised information, the organisation of which is counterposed to "the inquiry, dynamic, variable problem solving skills we expect". Mastery of factual knowledge, but lack of problem solving ability has also been noted by other workers (Mayou, 1978) and has been identified as creating difficulty in the transition from medical education to clinical practice (Farquhar et al, 1970; Knafl and Burkett, 1975). Pedagogical attention to the structural aspects and inter-relationships of aptitude and substance might diminish such divergence and difficulty. Ausubel et al (1978) discuss this in clear terms:

"The possession of relevant background knowledge (concepts, principles, transactional terms, "available functions") in cognitive structure particularly if clear, stable and discriminable, facilitates problem solving ... Without such knowledge, as a matter of fact, no problem solving is possible irrespective of the learner's degree of skill in discovery learning; without it he or she could not even begin to understand the nature of the problem confronted. Still another cognitive structure source of positive transfer inheres in applicable general elements of strategy, orientation, and set that reflect prior experience with related problems". (p. 572)

Thus Ausubel et al (1978) integrate in this statement the importance of structure, of substance and aptitude, their inter-dependence and the developmental role of experience. We may also infer from this that the problem solving skill, being

demonstrated in that inter-dependence and development, should most appropriately be learned where substance and aptitude coincide. Thus, it may be argued that cognitive skills training per se, taught separately from the clinical problems and then generalised or transferred to them (see section 4.2.4) might not be the most efficient and effective strategy for developing the diagnostic thinking process. It is already suggested in the earlier section cited that the indirect nature of the cognitive skills training method may have associated problems of structure and transfer.

On the basis of this conclusion, it is important to reassert that, despite criticisms of medical education's concentration upon acquisition of factual knowledge (Yonke, 1979), the role and importance of that knowledge must not be underestimated or even accorded less prominence than it deserves. Elstein et al (1979) also emphasise that reasoning strategies do not operate in the abstract but are applied in conjunction with the contents of memory. Ausubel et al (1978) are, again, particularly clear and forceful in their opinion. In discussing cognitive factors' in classroom learning they state that:

"Among these factors, the existing structure of knowledge at the time of learning (cognitive structure variables) is, perhaps, the most important consideration. Since this involves, by definition, the impact of all prior learning experience on current learning processes, it is co-extensive with the problem of transfer".

Witkin et al (1974) suggest reasons for this. They consider that amount of knowledge per se is not as important for cognitive clarity as is the degree of assimilation of knowledge. A small fund of facts may be appropriately assimilated or a large fund may not be appropriately assimilated. The latter circumstance will yield an account of them in the subject, when asked, which is confused, circumstantial, overspecific and overconcrete. This they describe as being of "out of focus" quality. Thus, students may learn factual information in the library or laboratory which is out of focus clinically. Consider, for example, the following textbook description:

"The term cyanosis is used to describe the clinical sign of a bluish colour of the skin and mucous membranes resulting from an increased amount of reduced hemoglobin, or of hemoglobin derivatives, in the small blood vessels of those areas". (Harrison, et al, 1971, p. 205)

Experience of cyanosed patients, however, may reveal that, although 'bluish' may be the most accurate word available in our vocabulary, it does not quite accurately describe the physical sign of cyanosis and students may require assistance in re-interpreting, or focussing, the meaning of the term in clinical practice.

Having thus argued for the importance of the role and position of knowledge or substance, we must now reconcile this position with our simultaneous contention that the diagnostic thinking process must be taught as a general skill or a range of possible skills and processes to be appropriately selected, used and noted during the clinical problem solving process. This latter conclusion reflects an acceptance of the tenet of Bruner (1973):

"Let me in general propose this test as a measure of the adequacy of any set of instructional propositions - that once they are grasped, they permit the maximum reconstruction of material unknown to the reconstructor ... General education does best to aim at being generic education, training men to be good guessers, stimulating the ability to go beyond the information given to probable reconstructions of other events".

This is reflected in the opinion of Barrows (1976) as he discusses the importance of equipping the medical student to be an effective thinker or problem solver:

"... the physician will never be able to choose the problem he wants to deal with. Patients always come to him as unknowns - unknowns in complexity, unknowns in urgency, unknowns as far as the systems involved - therefore we must add to the ability, that he has to deal with these problems as unknown problems".

How, then, is it possible to propose that the aim of medical education must be simultaneously generic and specific? This question is answered by reference to the structural aspects of teaching and learning. The instructional propositions of Bruner (1973) can be of substance, of aptitude or of the pedagogy itself. Each of these may be generic to the extent that the structural aspects are sound and appropriate. With regard to substance, one of the major conclusions of the two parallel studies is that the efficient and effective interpretation of clinical information is dependent upon the appropriate structure, content and, therefore, accessibility of the problem solver's store of knowledge. In addition, these same characteristics of the store have a determining influence upon the adequacy and completeness of the clinical enquiry in progress (see section 11.13). Thus, the specific aspect of the substance of medical education is, of course, the identification of content. But the generic aspect of substance is in the structure of its presentation and storage. Such structure will be powerful in determining whether or not the learner can apply his increased knowledge and skill to the next unknown problem.' This reflects Bruner's (1960) opinion quite closely:

"The best way to create interest in a subject is to render it worth knowing, which means to make the knowledge gained usable in one's thinking beyond the situation in which the learning has occurred". (p. 31)

This may be achieved, according to Bruner (1960), by structural means:

"... the curriculum of a subject should be determined by the most fundamental understanding that can be achieved of the underlying principles that give structure to the subject".
(p. 31)

In the current instance, of course, the structure of the subject concerns not only the characteristics of a body of knowledge, but also the characteristics of the cognitive processes of those who use that knowledge to solve problems. Thus, the structure of aptitude, which is identified as the

second type of instructional proposition of medical education, is also fundamental and may be seen as generic. Indeed, as some parts of the aptitude, such as the interpretation of clinical information by means of cognitive operations of extrapolation and structuring, appear to be common to all instances of clinical problem solving, its generic nature is readily perceived and must be accommodated by the pedagogy. However, the specific aspects of aptitude must also be identified. From the point of view of teaching and learning, these must concern the specific awarenesses and self-monitoring abilities which, it is proposed, medical education should encourage and facilitate in the learner in a conscious manner. It is suggested earlier in this section that detailed knowledge of such generic aspects of the aptitude as structure and extrapolation need not, necessarily, form part of the specific education of the clinical problem solver but that both students and practitioners should use the thinking and enquiry processes, which may be described in these terms, to greatest effect. We therefore must consider the third type of instructional proposition of medical education which is the pedagogy itself. The generic aspect of this may clearly be seen to be identified with its structure. How students learn, the contexts and methods of assimilation of new substance and aptitude, will determine whether or not present learning will serve the future. Having said this, there seems to be no opportunity for the pedagogy to have a specific aspect. It seems reasonable to suggest that the pedagogy itself is the key to learning and to development of the diagnostic thinking process.

We have considered substance and aptitude both in their specific and generic manifestations but all must be integrated and stored appropriately. The pedagogy which, eventually, is the major external determinant of the context and structure of learning is, by implication, the major external promoter of such integration and appropriate storage. Crutchfield (1972) is particularly lucid on this point and is worth quoting at length since he draws together much of the discussion of this

entire section. We may take his "master thinking skill" as the diagnostic thinking process:

"The ability of the individual to manage all these diverse and disparate requirements in the problem solving process has been termed the master thinking skill. This is a metaskill which enables the effective co-ordination, integration, and utilization of the many specific skills we have enumerated. Without this overall master thinking skill, an individual may be able to accomplish some parts of the process, but not others ... The educational implications of these comments about the master thinking skill are clear. If we seek to nurture the student's ability to think, then we must give him appropriate training on the many specific skills we have described. But in order to do this most effectively - so as to practise these skills in a mutually reinforcing way and to make for optimal transfer - we should train them simultaneously in the context of whole problems which have considerable scope, complexity and meaningfulness. In this fashion, the student will practise using his productive mental processes in the integrated way they must be used for genuine problem solving". (p. 196)

We may, therefore, suggest a tentative description of the type of learning which should be aimed for in relation to the diagnostic thinking process. Such learning should ensure the development of the applied diagnostic thinking process in practice on the basis of the structurally appropriate assimilation of an increasing body of relevant substance and aptitude. Before concluding with a statement of suggested aims in the design of teaching strategies, it may be helpful to summarise the arguments in the following points:

1. Specific sequences of diagnostic thinking processes cannot be prescribed and learned.
2. Undergraduate medical education should facilitate an awareness of the range of potential cognitive processes at appropriate levels of specificity and generalisability.
3. Teaching strategies must present the diagnostic thinking process as an entity amenable to conscious awareness and monitoring but not necessarily to particular shaping other than post hoc correction.

4. The nature and development of the cognitive structure of elements and processes are the ultimate determinant of the efficiency and effectiveness of the diagnostic thinking process. Therefore teaching and learning strategies must be devised on the basis of structural considerations.
5. Divergence and difficulty in the transition from medical education to clinical practice might diminish with greater pedagogical attention to structural aspects, inter-dependence and inter-relationships between substance and aptitude, and the developmental role of experience.
6. Clinical problem solving skill should be learned where substance and aptitude coincide and not as a separate cognitive skill.
7. The importance of the acquisition of factual knowledge must not be underestimated but that acquisition must be structurally appropriate and in the context of relevant aptitudes.
8. The design of teaching strategies must concern the specificities and generalities of substance and aptitude. However, the pedagogy itself may be identified as only having a generic aspect. It is the major promoter of the necessary integration of substance and aptitude and the development of appropriate storage structures.
9. Learning in relation to the diagnostic thinking process must be based on the structurally appropriate assimilation of an increasing body of relevant substance and aptitude.

In conclusion, we may propose aims in the design of teaching strategies as follows:

Aim 1: To facilitate an awareness of the range of possible cognitive operations in the diagnostic thinking process as defined in relation to its three stages (section 12.2.4).

- Aim 2: To facilitate self awareness and self monitoring during the clinical enquiry and problem solving process and the ability to compensate for or correct errors.
- Aim 3: To facilitate appropriate reciprocal development of the substance and aptitude of the diagnostic thinking process.
- Aim 4: To facilitate development of a foundation of substance and aptitude that is amenable to continued development during subsequent experiences of clinical practice and problem solving, and formal postgraduate education.

13.3 Pedagogical Approaches

The multiplicity of theories of learning (see, for example, Cross, 1974 or Hilgard, 1964), dearth of theories of teaching (Gage, 1964), and variety of viewpoints concerning curriculum development (Harris et al, 1975; Hooper, 1971; Stenhouse, 1975) make it advisable that the instructional designer should indicate his own approach or guiding principles. The following section, therefore, presents some discussion of the present background considerations and general orientation in the light of which possible pedagogical approaches are suggested. Subsequent sections concern aspects of the pedagogy itself.

13.3.1 Contextual Considerations

Clearly, a major determinant of suitable pedagogical approaches in development of the diagnostic thinking process is the psychological interpretation of the process itself. Thus, the structuralist theory which is developed in Chapters Eleven and Twelve to describe and explain the clinical problem solver's cognitive processes and characteristics, is reflected in the associated pedagogical orientation pre-saged in Chapter Four.

However, other types of consideration are also influential. We may begin with one which is purely pragmatic. In identifying strategies to facilitate development of the diagnostic thinking process, it must be borne in mind that the discussion is in relation to British medical schools which have the opportunity for educational innovation within certain limits of finance, staffing levels, patient availability, the necessary core curriculum, a particular examination structure and so on. In addition, and perhaps even more important than these factors, are the teachers with their existing views and experiences of their roles and responsibilities, and the students with their habits and views of learning. While the pedagogy must be designed to facilitate learning, the crucial role of the facilitator must not be overlooked. Bruner (1960) makes this quite clear:

"A curriculum is more for teachers than it is for pupils. If it cannot change, move, perturb, inform teachers it will have no effect on those whom they teach. It must be first and foremost a curriculum for teachers. If it has any effect on pupils, it will have it by virtue of having had an effect on teachers. The doctrine that a well-wrought curriculum is a way of 'teacher-proofing' a body of knowledge in order to get it to the student uncontaminated is nonsense".

Any proposed pedagogical approach, then, must take into account these contextual factors and be practical and realistic within their constraints. These, however, are not the only considerations which play a part in shaping the eventual teaching and learning strategy.

Any approach to course design must rest on certain implicit or explicit assumptions about the learners for whom the course is intended, about the most appropriate framework for presentation of the course material, and about the possible or probable forms of interaction between learner and course material. With regard to each of these, our current assumptions are that the learners apply some subset of their general range of cognitive processes to the diagnostic problem and would do so regardless of the pedagogy to which they are exposed. However, it is also assumed

that the pedagogy can either enhance or hinder the development of an efficient, effective and appropriate selection and application of that subset of operations. Further, the assumption is made that no cognitive prescription can be made to guide the thinking processes of the clinical problem solver in advance of his problem solving activity, but that his education may equip him to generate and apply his own means of cognitive self monitoring and control when the occasion arises, as it will many times daily during his future clinical practice. On the basis of these considerations, it is concluded that the presentation of the substance of undergraduate medical education must have due regard to the developing and mature cognitive processes and structures associated with its use in clinical practice and problem solving. These assumptions and conclusions reflect the approach to course design of Mace (1976), which she describes as follows:

"It assumes that the University ... bases its courses on one (or a combination) of the forms of knowledge, or disciplines, which represent our past and current ways of trying to make sense of the world. When he engages in study of them a student becomes actively involved in the attempt to create such an understanding for himself. He must be introduced to the central concepts (or abstract generalisations), procedures (methods by which discoveries are made), and criteria (principles according to which 'proof', or what counts as 'evidence', is established), that go to make up the subjects under study and that will enable him to think for himself within them. Moreover, I am assuming that the way in which that introduction is made will determine the nature and range of his thinking". (p. 26)

Thus, Mace concisely describes the approach, assumptions and areas of substance and process which are adopted in the present discussion. She also indicates very clearly the set of inter-relationships between substance, process, teaching and learning. Her view of the learner also is redolent of the present study which recognises the importance of the student's already formed and forming cognitive structures and processes. From this vantage point medical education may be seen as being

yet another, albeit special, formative influence or experience which yields the cognitive assimilations and accommodations which all other experiences also occasion. Thus, it is important to recognise that medical education, like any other, merges with a stream of cognitive development and, while not necessarily affecting its flow, may alter its course. Mace (1976) summarises this assumption also when describing her view of the learner:

"... he is a person with purposes, concerns and interests of his own ... Although he understands what is being offered in terms of his own purposes, concerns and habits of thought, in the very attempt to understand, those purposes, concerns and habits of thought are themselves modified and transformed. When he acts selectively in this way on what is being offered he also, in a sense, transforms that. In short, in the attempt to create his own understanding he is both acted upon and acting upon".
(p. 26)

The present structuralist approach is very much in accord with this opinion. Such an approach has an inevitable implication for course design. Mace (1976) argues that attempts to define closely the precise intended outcome of learning fail to respect the attributes of the learner which she identifies, and the dynamic relationship between learner and course material which may be inferred. The general and broad nature of the aims proposed for the design of teaching strategies in relation to development of the diagnostic thinking process (section 13.2 conclusion) reflect Mace's concern. Each of these four aims suggests the implicit assumption that the learner must be enabled to define for himself his own interpretative responses, perceptions, structures and processes and that, in the end, substance and aptitude will be assimilated and developed on the basis of the learner's previous learning and development of aptitude and substance. It is for these reasons that awareness, not prescription, is advocated.

We may now follow Mace's (1976) argument through to its conclusion which is in accord with that of Stenhouse (1975):

"We can take the structure of the subject in question as our touchstone. Our aim would be the clear and honest explication of the sorts of procedures, concepts and criteria that are at work, and this aim would inform our selection of material for the course. In our attempt to be clear about the structure of the subject, ways of presenting it to others would suggest themselves consistent with our view of those others as purposeful and interested". (p. 26)

For medical education, structure here must refer both to the relationship between substance and aptitude and to the relationship between learner and that which is to be learned, as well as to the structural aspects of these elements themselves.

Having considered some assumptions which underlie our selection of possible pedagogical approaches in development of the diagnostic thinking process, we may now turn to another, related contextual consideration. This concerns our knowledge of cognitive psychology and its implications for education. A highly pertinent discussion is presented by Broadbent (1975) who draws some conclusions apposite to the present frame of reference and reinforcing to the previous assumptions and conclusions. With regard to the relationship between formal education and the stream of cognitive development of the learner, Broadbent (1975) (considering child rather than adult learners) points out that:

"By the time a teacher meets any child, some system of organisation will probably already exist, and it is likely that fresh information will be best assimilated through the system which is already tentatively established".
(p. 175)

Our present concern with the cognitive autonomy of the student, or the primacy of the student's own cognitive structure at the time of learning, is thus reinforced. However, Broadbent (1975) also draws attention to the possible inhibiting effects of that cognitive structure in terms of the resultant selectivity of the learner's attention. Such selectivity might be set by the learner's previous learning.

"Only some of the information presented will receive attention, and if this is not decided deliberately it will certainly be decided by chance factors". (ibid)

Such a phenomenon is noted in the account gathering study as the set effect (section 11.9.1). It may apply equally to the type of formal learning now under consideration. The matter is also discussed in 4.1.1 above, where it is pointed out that regardless of the objective content of the environment, where the learner is left unguided his own prior knowledge will determine which parts of that environment will function as a stimulus for him. The teacher's role, therefore, must include having due regard to the context of the student's prior knowledge structure and ensure that all the elements of the clinical or instructional environment are attended to and that new or incongruent ones become structured in the learner's memory store appropriately (or old ones become restructured), and used to advantage in his developing problem solving strategies. This conclusion reflects two more of Broadbent's (1975) messages from cognitive psychology to education:

"When something is noticed in the environment, it can be processed or encoded in a number of ways, and the particular processing which takes place will decide whether the effects are long lasting or transitory".

"If the effects are to be enduring, then the information must be organised at the time it is stored, in such a way that a clear path leads to it from the likely situations in which it may be needed".
(ibid)

It is clear that this area of our discussion may be summarised as the relationship between the development of substance and aptitude on the one hand and the teaching and learning process on the other. The views of Bruner and Gagné on this relationship are usually interpreted as antithetical (Shulman, 1972), yet they seem quite reconcilable and consistent within the present context. Firstly, consider Bruner's (1966) view. His discussion happens to be in relation to mathematics teaching:

"... a theory of instruction seeks to take account of the fact that a curriculum reflects not only the nature of knowledge itself - the specific capabilities - but also the nature of the knower and of the knowledge - getting process. It is the enterprise par excellence where the line between the subject matter and the method grows necessarily indistinct ... We teach a subject, not to produce little living libraries from that subject, but rather to get a student to think mathematically for himself, to consider matters as a historian does, to take part in the process of knowledge-getting. Knowing is a process, not a product". (p. 72)

The structuralist approach which is here adopted with regard to the process of medical education, reflecting the structuralist interpretation of the diagnostic thinking process itself, reflects Bruner's view in its recognition of the close and necessary inter-relationship between substance and aptitude. Substance must be acquired in a manner and to an extent consistent with and conducive to the development of efficient and effective aptitude. The point of apparent reconciliation between the positions of Bruner and Gagné, however, lies in this very necessity to acquire substance, and in the manner of doing so. Gagné (1965) agrees that formal instruction should have the aim of "teaching the student how to think" (p. 170) but doubts that this can be done independently of content. With regard to such strategies or styles of thinking, Gagné (1965) considers that:

"Even if these could be taught (and it is possible that they could), they would not provide the individual with the basic firmament of thought, which is subject-matter knowledge. Knowing a set of strategies is not all that is required for thinking; it is not even a substantial part of what is needed. To be an effective problem solver, the individual must somehow have acquired masses of structurally organised knowledge. Such knowledge is made up of content principles, not heuristic ones". (p. 170)

Although it is undeniable that knowledge is made up of content principles, Gagné seems to take insufficient account, firstly, of the structural organisation to which he draws attention, and secondly, of the possibility that knowledge can be of

process as well as substance. In this light, although knowledge may be made up of content principles not heuristic ones, it is equally undeniable that the former surely serves the latter, and in some instances (such as development of concurrent self awareness of the diagnostic thinking process) the former and latter may be identical. With regard to the pedagogy of the diagnostic thinking process, then, the approaches of Bruner and Gagné are mutually enhancing rather than mutually exclusive.

Before progressing to a more specific discussion of that pedagogy, it might be useful to consider other aspects of the contextual factor of the relationship between manner and outcome of learning. Stones (1966) takes a similar approach to the interpretation of learning as does the present discussion. We may understand his schemas to be as our structures:

"Thus learning advances by a series of schemas which increase in complexity as they assimilate new elements. The use of cognitive schemas may be regarded as an important aspect of meaningful learning. Learning which does not make use of such schemas will have less wide application. Behaviour which does not involve schemas will tend to consist of arbitrary linkages and responses. This we shall refer to as rote learning". (p. 147)

Thus, as Osgood (1953) points out in relation to Wertheimer's work, "we cannot expect forceful, productive thinking in problem situations from people who are trained by blind rote methods". However, it is possible and useful to consider the question of the relationship between style and outcome of learning from a different viewpoint. So far, we have considered only the effects of learning both in terms of memory store and problem solving ability resulting from use of that store. We have considered the knowledge-getting process mentioned by Bruner, only from the point of view of future effects of present learning. However, we may also consider the process of knowledge acquisition from the point of view of the approach adopted by the learner.

Marton and Säljö (1976) identify two levels of information processing or approaches to learning in university students. These levels they call deep level and surface level processing and describe in the following terms:

"In the case of surface level processing the student directs his attention towards learning the text itself (the sign), i.e. he has a reproductive conception of learning which means that he is more or less forced to keep to a rote learning strategy. In the case of deep level processing, on the other hand, the student is directed towards the intentional content of the learning material (what is signified), i.e. he is directed towards comprehending what the author wants to say about, for instance, a certain scientific problem or principle". (pp 7-8)

Transposing this into the current framework, and in relation to development of the diagnostic thinking process, to grasp the significance of new material is to relate it to previous learning and to perceive its inner structure. For Marton (1975) "to grasp what is signified is simply to discover (or to create) meaning. In our opinion, this is precisely "what it takes to learn". And - we may add - to teach is to facilitate this learning". Given the nature of problem solving as one type of learning, we may reasonably say that the task of the medical teacher must be to facilitate something akin to Marton's deep level processing. Marton and Säljö (1976) are in no doubt that such a level of processing is amenable to facilitation. Their 1976 study shows that students can adapt their way of learning to their conception of what is required of them. The present discussion reflects their conclusions:

"The fundamental importance of recognising the necessary link between the level of processing adopted by the student and the level of understanding cannot be overstated. Students adopt an approach determined by their expectations of what is required of them. While many students are apparently capable of using deep or surface strategies, it may be that the

current demands of the examination system at school level are interpreted by them as requiring mainly the recall of factual information to the detriment of a deeper level of understanding. The present investigation suggests that students may need to refocus their attention on the underlying meaning of what they are required to study and that this process could be helped by ensuring that the assessment procedures demand deep level processing. However, the existence of a group of students who 'technified' their attempts at deep level processing indicate that a more prolonged, and perhaps more explicit, redirection of attention may be necessary". (p. 125).

Howe and Colley (1976) add further experimental support to the contention that the nature of learning is influenced by the set created by the nature of questions put to the learner following previous learning. This certainly has implications for pedagogical as well as assessment procedures. Whether the set for learning is created by the examination system alone is a matter for debate. It would seem reasonable to contend that it is also created, in the short term, by the pedagogy itself. If the four aims defined earlier (section 13.2) in relation to development of the diagnostic thinking process are adopted, then appropriate assessment procedures must ensue. In the meanwhile, the present discussion addresses itself only to the pedagogy and not to assessment strategy. The central issue concerns the means and methods whereby the deep level processing associated with appreciation of underlying meaning and comprehension of what is signified may be facilitated both in relation to substance and aptitude, to the content and process of preparation for clinical practice. Such a pedagogy may be determined in relation to our knowledge of the developing and experienced diagnostic thinking process and to our inferences concerning the structure of substance and application of aptitude and the relationships between these elements.

These, then, are the preliminary contextual considerations to be noted in any discussion of the pedagogy of the diagnostic thinking process. They may be summarised as follows:

1. Psychological nature of the developing and experienced diagnostic thinking process.
2. Inferences concerning the substance and aptitude of the diagnostic thinking process and their inter-dependence and inter-relationship.
3. Limiting features of the medical school environment and curriculum.
4. Present approaches and roles of medical teachers and teaching.
5. Assumption of the learner's necessary and inevitable application of a subset of the general range of cognitive processes.
6. Assumption of the formative effect of the pedagogy.
7. Assumption of the inappropriateness of precise prescription of a diagnostic thinking process, and of the possibility of equipping the student to apply cognitive self monitoring and control of his own contemporaneous diagnostic thinking process.
8. Inter-relatedness of substance, process, teaching and learning.
9. Recognition of the importance of the learner's inchoate, forming and formed cognitive structures and processes in facilitating or inhibiting learning.
10. The implications of cognitive psychology for education in terms of selective attention and manner of the learner's assimilative and accomodatory responses.
11. Recognition that the pedagogy must reflect the natures of the substance and the aptitude.
12. Recognition of the deterministic realtionship between style and outcome of learning.

Having considered the context of the pedagogy, it may be most logical to progress to a consideration of its content.

13.3.2 Curriculum Content and the Role of the Teacher

The question of content is inseparable from that of the role of the teacher, since what is suggested is not any radical addition or subtraction from the content of any undergraduate curriculum but rather an alteration in the presentation of that content. It is argued above (section 13.2) that the separate study of the diagnostic thinking process would be inappropriate and, probably, counter-productive. The defined aims (section 13.2) do not concern the student's learning of any theory. They concern, firstly, the development of habits of self monitoring and self awareness. Secondly, they concern the reciprocal development of substance and aptitude such that the formed cognitive structures and processes will serve the future in the manner described in Aim 4 (section 13.2). The first of these involves, of necessity, some knowledge of the potential cognitive processes. But it is suggested that such knowledge should be inferred from and be based on or grow out of the learners' practice of clinical problem solving. The second of these, however, necessitates that the teacher has a sound knowledge of the range of possible cognitive processes which will be demonstrated by his students and that he has the pedagogical (and analytical) skills to guide each student to a practical and useful understanding and awareness of his own cognitive processes. In addition, it necessitates a basic knowledge and understanding of the structuralist rationale. It is emphasised that such knowledge of theory need not be highly elaborated in order to be useful.

Wood et al (1976) discuss the role of tutoring in problem solving. The tutorial process they conceive as an "interactive, instructional relationship" whereby an expert helps someone who is less expert. Wood et al (1976) are considering only the situation in which a student is assisted in solving a problem by having the teacher control the elements of the

task that are initially beyond the learner's capacity, thus permitting him to concentrate on and complete only those elements that are within his range of competence. Although this may be part of the medical teacher's proposed task, he has the additional one of guiding the student to an awareness of the way in which he is approaching the clinical problem and deciding upon a resolution or diagnosis. We are, perhaps, echoing Bartlett's (1958) opinion that thinking is a form of skill, that it has acknowledged experts and that it has to be acquired by "well-informed practice". The teacher must be acknowledged expert and informer in one. But Ausubel et al (1978), using an apposite analogy, make it clear that this dual role is unlikely to occur without preparation:

"Teaching, like medicine, requires a long period of practical apprenticeship as well as particular sensitivities, diagnostic skills and the ability to prescribe and implement suitable practices. These competencies go beyond what is learned in applied science courses such as educational psychology and child development (or medical physiology and pathology), or in clinical courses such as methodology and student teaching (or clinical diagnosis and clinical clerkship). Also, the individual skilled in appraising an educational situation (or in diagnosing a patient's condition) is not necessarily equally skilled in proposing and putting into practice effective measures for learning (or treatment). But judgment without knowledge of principles is no more effective than knowledge of principles without judgment. Neither is it any more likely to be free of error".
(p. 8)

In their discussion of the necessity for preparation of the teacher, Ausubel et al (1978) return us to the question of content. Given that it is the pedagogy of a process that is under consideration, definition of content may seem somewhat incongruous. However, given also that the structure and elements of that process are defined and described, these must (at some appropriate level of specificity) determine or constitute the content of the instructional process. Thus, the teacher would find it useful to have at his dis-

posal a basic knowledge of the fundamental psychological features of structure and extrapolation accompanied by more specific (although, again, not necessarily highly elaborated) knowledge of the major characteristics of the three stages of the diagnostic thinking process as summarised in section 12.2.4 above. Such knowledge should enable the teacher to analyse and interpret the thinking processes of the student and to guide him towards his own understanding and knowledge in turn. Complementarily, the structure or process of the pedagogy must reflect and enhance the structure and course of the diagnostic thinking or problem solving process itself. However, achievement of such aims is equally dependent upon the design and implementation of teaching strategies which expose the learner's diagnostic thinking process or make it otherwise amenable to interpretation and guidance or remedy. The nature of such teaching strategies is the subject of the following two sections. In preparation, we may summarise the present section in the following points:

1. No radical alteration of curriculum content is proposed, but alteration in the presentation of content is advocated.
2. The medical teacher should have a sound knowledge of the range of possible cognitive processes and the fundamental features of structure and extrapolation.
3. The medical teacher should have the pedagogical (and analytical) skills to guide the student to practical and useful understanding and awareness of his own contemporaneous diagnostic thinking processes.
4. The defined structure, elements and explanatory concepts of the diagnostic thinking process should determine or constitute the content of the associated pedagogy.
5. The structure or process of the pedagogy should reflect and enhance the structure and course of the diagnostic thinking process itself.

We may conclude by placing the curriculum content and the role of the teacher in their wider context:

The pedagogy, which is implemented by the teacher, should reflect a structurally appropriate increasing store of knowledge and a developing applied thinking process.

The teacher should guide the student towards self awareness, monitoring and remedy of the contemporaneous diagnostic thinking process.

Prospectively, these should prepare for the ready future development of substance and aptitude by means of two complementary aspects:

A process of expedient teaching strategies.

A content of a serviceable analytical and synthetical framework of theoretical constructs and defined processes as summarised in relation to the three stages of the diagnostic thinking process.

13.3.3 The Pedagogy of Structure

It is argued that pedagogy in the development of the diagnostic thinking process must reflect the structural characteristics of stored knowledge, as well as enhancing the development and use of such knowledge in the process itself. The present section considers pedagogical implications of structure and therefore of substance, while section 13.3.4 concerns the pedagogy of the cognitive process itself and therefore of aptitude.

Armstrong et al (1979), in the context of interviewing, diagnostic procedures and the doctor-patient relationship, point out the importance of the internal representation of the problem for the solver, "since the way in which a problem is represented and defined will largely determine how it will be tackled". The discussion of structure is likewise one of internal representation which is partially dependent upon the context and manner of initial presentation of information.

However, it would be erroneous to assume that information presented during some real or simulated process of clinical problem solving is also automatically stored within that context. It is concluded earlier (section 4.2.2), for example, that problem based learning leaves the question of structure open. The organising agents in problem based learning are not identified, thus no conclusions may be drawn about the structural properties of knowledge acquired through such learning. In turn, it cannot be assumed with any justification that such stored information will display greater ease of retrieval or transfer in the face of a new clinical problem. Unless the characteristics of the problem solver's own thinking are known and taken into account, structural generalisation (and thus transferability) is difficult to achieve by design. With regard to the development of the diagnostic thinking process, pedagogy should be designed to make full use of and maximally enhance the cognitive structure of the learner if it is known how he uses that structure in his thinking. The present study provides clear indicators of this process. Section 12.2.4 summarises the main characteristics of the three stages of the diagnostic thinking process. Of these, certain features may be considered as of special importance in any discussion of the pedagogy of structure in development of the diagnostic thinking process. These features concern the structural characteristics of stored information and the structural characteristics attributed to the changing array of information which constitutes the clinical problem. We may identify them as:

<u>Stage One</u>	Differential forcefulness of items of information.
	Attribution of force by the selecting operation.
	Absolute threshold of force.
	Access to stored structures from any item of information.
	Varying degrees of appropriateness of extrapolated contexts.

Stage Two Restructuring by shift of force.

Restructuring by multiple extrapolation from
the same forceful feature.

Stage Three None.

From these features, it may reasonably be concluded that the pedagogy of structure involves learning not only the information itself, but also its possible interpretative values in relation to other items of information, either presented or stored. It involves also learning that the same information can be perceived in different ways depending upon its context (actual or extrapolated) or upon the relative importance accorded to it. In brief, it involves learning that information can be cognitively re-arranged and acquire new meanings and significance, just as pieces on a chess board can be physically re-arranged and, because of the structure of rules of the game, also acquire new meanings and significance. The pedagogical questions, however, are How? and When?

In section 4.2.2 it is asked: What are the organising agents in problem based learning? Subsequent discussion enables us to answer that. They are the problem solver's identification of forceful features and his methods of arranging information in stored structures according to his (more or less accurate and complete) perception of his own clinical problem solving experiences. These organising agents may be limited or distorted by previous experiences of learning outside the clinical problem solving context. The quality of information being 'out of focus' (see section 13.2) may generalise by pro-active interference to new information. Having said this, however, it must also be recognised that a process of retro-active interference could just as well bring such information sharply into clinical focus. Thus an experienced clinician understands very clearly the meaning of 'bluish' when used to describe the cyanosed state, unlike the clinical inexperienced student who, having read the term, may be excused for seeking a patient of cobalt or ultramarine hue. Such a line of

argument leads us inexorably to question whether or not there may be occasions when it is not actually necessary or most profitable for a medical student to acquire information in a structure appropriate to his clinical problem solving thinking process. After all, no cognitive structure is inviolable, none is immune from restructuring and some information, perhaps, may best be initially understood and assimilated within a context other than that of the diagnostic problem solving process. We may think, for example, of microbiology, chemical pathology, physiology or anatomy. Within the purview of the present discussion, it is necessary only to answer that the same information may find itself located in more than one structure (as multiple extrapolation from the same item or items of clinical information indicates) and that a necessary pre-requisite of restructuring is an ability to identify the elements of the field of information (see section 12.2.1; Dickstein, 1968; Witkin et al, 1974).

The fourth of our identified aims (section 13.2) concerns the development of a foundation of substance and aptitude that is amenable to continued development. Such an aim must also apply to any station of the undergraduate curriculum, as well as to its endpoint. Giving pedagogical consideration to structure thus requires attention to the possibility of a range of sequential structures. Vertical integration of the curriculum may seem an appropriate environment in which to achieve exactly this. Without it, premature acquisition of inappropriate, unalterable structures may result in premature closure which inhibits the acquisition of more appropriate structures later (Ausubel et al, 1978).

We have identified the learner's organising agents in the clinical problem solving process as his identification of forceful features and his arrangement of information according to his perception of the field and extrapolation to stored embedding contexts. The pedagogy of structure must aim to achieve a conscious and clear use of these agents. It is

suggested that they will operate regardless of the learning environment. The pedagogical task is appropriate arrangement of that environment to facilitate such operation to maximal efficiency and effect. Ausubel et al (1978) advocate the use of advance organisers in learning 'to enhance proactive facilitation' (see 12.2.1 above). In the present context, it is suggested that the whole of undergraduate medical education may be seen as the advance organiser for clinical practice and clinical problem solving. Further, it is suggested that each learner must develop his own structures and processes through an awareness of his own cognitive operations. The pedagogy of structure may yield the advance organisers, but must itself be a concurrent organiser.

Some clarity in defining the role and psychological meaning of structure in the diagnostic thinking process has been achieved. To take advantage of this, the pedagogy may most appropriately encourage the student to rehearse actively and consciously the elements of that process which uses stored knowledge and to involve himself in the equally active identification, organisation and re-organisation of the elements of the field as they are elicited during the clinical encounter. This must be done in relation to the stored structures in order to achieve rational and useful acquisition of knowledge. For example, the process, means and mechanism of extrapolation may be demonstrated by the student himself in ways to be described below and an awareness established of the cognitive processes which usually remain hidden from analytical view and are often represented as intuition. The basis of such a strategy, it is suggested, must be the identification of the elements of information available to the problem solver and the opportunity for subsequent application of active and conscious cognitive operations. Let us consider the rationale for this statement.

We have already noted (section 4.1.1) Lunzer's (1968) and Broadbent's (1975) illumination of the dilemma whereby what functions as a stimulus for the learner is not necessarily an

objective reflection of the field but may be determined by his prior knowledge (or schemata or structures). Thus for new structures to develop or established ones to become appropriately restructured, it may be helpful in the first instance to guide the learner's perception towards recognition or identification of the elements of the clinical situation, including those non standard ones identified in section 11.10 above. When he has perceived the field, the student might be encouraged to classify and rehearse its structural properties and the means of assigning them. For example, extrapolated contexts might be identified on the basis of the elements, groups of elements may be identified and interpreted and expectations discussed. In relation to the selecting operation, consideration of the relative force or value of items in yielding possible extrapolated contexts might be undertaken and reasons for differential forcefulness considered in relation to the learner's own stored structures, the clarity of information elicited and so on. The degree of satisfaction or cognitive ease the problem solver experiences in working with different types, items and arrays of information may be made explicit and analysed. The role of the teacher in all this might be to guide the learner towards satisfactory and serviceable interpretation and understanding of the relationship between items of information elicited and that stored in his own memory and to add to or restructure that store appropriately to future practice. Relationships, conjunctions, disjunctions, intersections and so on may be pinpointed so that the structural properties of new and established information may be clarified and addition of the new to the store may thus be more structurally sound than it otherwise would be. A discussion of specific teaching methods or styles is not intended here, yet the potential for both tutorial and group work as well as for individualised learning is clear. Likewise, the opportunity for integration of prior and present learning is also presented. A further product of such an approach may be to provide the problem solver with means of changing course of identifying new avenues to explore during the clinical interview when as happens (Armstrong et al, 1979)

he runs out of questions after, presumably, having run out of extrapolated contexts or items within them. The ability to separate the field into its component parts and evaluate each of these may also encourage avoidance of the set effect (see 12.2.2. above) by enabling more rational comparison of observed and expected findings.

Barrows (1973) suggests that groups or individual students should approach problem solving by asking such questions as:

"Is there a problem here, is there more than one?
Can the problem or problems be defined?
What data or observations are needed to more
accurately characterise the problem or problems?"
(p. 60)

However, such an approach does not reflect the basis of the diagnostic thinking process in cognitive operation on the elements or a subset of the elements of the clinical situation as it progresses and develops. Neither does it reflect the structural aspects of the cognitive processes of the clinical problem solver. The present proposed pedagogical approach relies on the fundamental elements and aptitudes of the diagnostic thinking process and upon making these amenable to conscious awareness and use. In addition, and perhaps more importantly, the proposed approach, which is based on the interpretation and structured storage of the substance of the clinical problem (the symptoms, signs and details of the clinical history) would appear to facilitate transfer of learning. Such transfer may be of aptitude and substance, from problem to problem, or from time of acquisition of what has been learned if such learning occurred outside the context of a clinical problem, to the time of its retrieval in order to solve a clinical problem or to be related to new learning.

The proposed strategy has the further and important characteristic of allowing for individual differences in the content of cognitive structure by allowing each learner to take advantage of and work from his own present state. The overt nature of

the types of teaching method indicated also allows the teacher some access to the student's cognitive structure by receiving specific information about the connections and extrapolations which the student makes on the basis of items or arrays of information presented. This reflects Lawless's (1979) opinion:

"Awareness of the importance of making connections between new and existing knowledge is the important quality. Devices such as 'advance organisers' (Ausubel, 1978) which acquaint the learner with the overall structure of the subject matter, and developing an overview by a 'spiral curriculum' (Bruner, 1966), have been shown to be effective, but the long term aim must be for the student to develop techniques of working out structure for himself".

Further, if indirect, support is given to the proposed information based approach if we reconsider the results of the present questionnaire study in relation to the philosophy of neurological diagnosis and teaching. The study demonstrates the similarity of students and registrars in their reliance on the tested and untested areas of learning and thinking necessary in clinical problem solving. Neurological teaching tends to stress "the need for a full history and a full description of all findings" (GMC, 1977). Thus the neurological emphasis on the information itself appears to yield in students cognitive characteristics more near those of the expert than is the case in endocrinology which has no such explicit emphasis. However, this statement must be tempered by recognition of the relatively less frequent success in actually making the correct diagnosis in neurology than in endocrinology, when students and registrars are compared. This may indicate that explicit attention to the elements of the field of information may assist development of process but not of helpful structures unless the structural characteristics and the inter-relationships between new and stored items of information are explored in an equally explicit manner. Broadbent (1975) stresses this point, that even when a learner "has noticed something about a situation, the way

it has been noticed is important". To support this contention, he cites Morton's (1967) work which showed that most regular telephone users in his study could not reproduce the positions of letters and digits on the telephone dial. Broadbent (1975) concludes that:

"Cognitive psychology not only warns us that information we present may fail to get into the nervous system at all, but also makes it clear that the future effects of what is seen or heard depend very much on the categories or codes into which information is placed ... The storage of information is much less likely to cause trouble if certain steps are taken at the time of storage".

It is suggested that the proposed teaching strategy pays due attention to these points by focussing attention on the field of information items and making it possible to clarify the student's stored structures in relation to these. Such guided but active, and potentially individualised learning, provides the opportunity to use the structured store, to restructure or otherwise alter or refine it, and to draw attention to the related cognitive process aspects of clinical problem solving. Cutler's (1979) clue orientated approach to teaching "the transformation of a data base into a problem list" perhaps approaches the present stress on the perception of information but lacks the necessary accompanying awareness of the complementary cognitive process.

It is appropriate now to consider the pedagogy of process but before doing so we may summarise the discussion of the pedagogy of structure:

1. Pedagogy in the development of the diagnostic thinking process must reflect the structural characteristics of stored knowledge and enhance its development and use.
2. The pedagogy of structure involves learning information (factual knowledge) and its interpretative value in relation to other presented or stored items of information. In addition, it involves learning that information can be

cognitively rearranged and thereby acquire new meanings and significance.

3. Stored structures are subject to both pro-active and retro-active interference.
4. A necessary pre-requisite in the pedagogy of structure is facilitation of the ability to identify the elements of the field of information.
5. The pedagogy of structure must aim to achieve conscious and clear use of the clinical problem solving processes of identification of forceful features and arrangement of information according to perception of the field and extrapolation to stored embedding contexts.
6. Undergraduate medical education may be seen as the advance organiser for clinical practice and problem solving.
7. The learner must develop his own structures and processes through an awareness of his own cognitive operations.
8. To achieve rational and useful acquisition of knowledge, the student should rehearse actively and consciously the elements of the process which uses the stored structures, and should actively identify, organise and reorganise the elements of the field elicited during the clinical encounter.
9. Teaching methods might be designed to guide the student's perception towards identification of the items of information in the field, his means of assigning structure by extrapolation, the structural properties of items of information, their potential for restructuring or reorganisation, selection of forceful features, attribution of relative force, satisfactoriness of extrapolated contexts and so on.
10. The role of the teacher must be to guide the perception and understanding of the learner.

11. Ability to analyse and evaluate the field of information may encourage avoidance of the set effect.
12. It is suggested that a structurally based pedagogy reflects the learning and thinking processes of the clinical problem solver; that it facilitates transfer of learning; and that it allows for and uses to advantage individual differences in the content of cognitive structure.
13. The proposed teaching strategy allows the teacher some access to the student's cognitive structure.

13.3.4 The Pedagogy of Process and its Relationship to the Pedagogy of Structure

The discussion of the pedagogy of structure closes with an indication of the essential unity or inter-relatedness of structure and process in the development of clinical problem solving skill. The present distinction between structure and process is, therefore, artificial and only useful for analytical purposes. The pedagogy of structure and the pedagogy of process in practice, promise to be the same activity, just as problem solving can be seen as a special type of learning (see section 4.1.3 above) and therefore has implications for the learner's structure of knowledge. However, just as being presented with information is shown in the previous section to be inadequate in ensuring the appropriate development of stored structures, so is practice in solving problems as such also inadequate without guidance and feedback about the cognitive processes used (see section 4.1.3 above).

It is suggested that the aims of the pedagogy of process are defined particularly as the first two of the four aims identified for the design of teaching strategies (see the conclusion of section 13.2 above). These concern the development of awareness of the range of possible cognitive operations of the three stages of the diagnostic thinking process and the facilitation of self awareness, self monitoring and, if necessary, remedial thinking. The remaining aims

concern the conjunction and outcome of teaching for both structure and process. The content of the teaching primarily must be derived from the cognitive operations identified in the account gathering study (see Chapter Eleven) and summarised in section 12.2.4. Thus the teacher should be able to guide the student towards an awareness, use and control of the cognitive operations of:

- | | |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <u>Stage One</u> | The selecting operation.

Extrapolation, single and multiple from single or many items. |
| <u>Stage Two</u> | Restructuring and re-interpretation by shift of force or multiple extrapolation.

Comparison of expected and observed findings by working from the extrapolated context.

Judgment of acquisition of sufficient information.

Reciprocal determinative and responsive cognitive operations and behaviour in relation to the flow of information.

Possible errors, e.g. set effect (see section 11.9 above). |
| <u>Stage Three</u> | Judgment of congruence of observed and expected information.

Possible reversibility of the judgment.

Methods of resolution of the problem. |

However, although these aspects may be identified as important for the student's awareness, their separate identities do not, while separate, add up to the diagnostic thinking process. Crutchfield (1972) expresses this point:

"... the thinker's purpose is to solve a problem by using many specific skills. These skills must be effectively co-ordinated with one another if the problem solving process is to go forward to a successful conclusion. The perfecting of one of these specific cognitive skills is thus as much a matter of learning how to integrate it with the other concurrently operating skills as it is a matter of practising this skill by itself. In short, what is required is the development and strengthening of the master thinking skill". (p. 196)

Full support is given to Crutchfield's viewpoint. Further, qualified support is also given to the pedagogical lessons which he draws from this conclusion:

"The educational implications of these comments about the master thinking skill are clear. If we seek to nurture the student's ability to think, then we must give him appropriate training on the many specific skills we have described. But in order to do this most effectively - so as to practise these skills in a mutually reinforcing way and to make for optimal transfer - we should train them simultaneously in the context of whole problems which have considerable scope, complexity and meaningfulness. In this fashion the student will practise using his productive mental processes in the integrated way they must be used for genuine problem solving". (p. 196)

From this, it is clear that our support for this position must be qualified simply because the aims of the pedagogies are different. It is our contention that the diagnostic thinking process is not special in its form, but only in its application. The diagnostic thinking process is, it is suggested (see Chapter Twelve) an everyday cognitive process. The skills are already integrated. The training is not in problem solving per se. The aim of the pedagogy of process must be to facilitate analysis (and therefore separation) of the component cognitive skills. It is not suggested that any new thinking process be taught or learned, but just that those which are applied are monitored and controlled and, if necessary, made less prone to error. Thus the training of the pedagogy of process is not like that of Crutchfield (1972). It is, rather, training in self awareness. The learning which is intended concerns the relative balance of the separate skills of the

diagnostic thinking process as described in Chapter Twelve.

Section 12.3 above discusses and summarises the developmental aspects of the diagnostic thinking process. It is shown that development is characterised by a change in balance and relative contribution of the various separate cognitive operations. No new operations appear to develop or be applied as a result of experience. Rather, a more efficient and effective use of those already extant is achieved. It is suggested that with appropriate pedagogy the experienced balance might be achieved more rationally and more efficiently than the current apparently trial and error method allows. Peel (1971) writes in terms of teaching being directed at enabling thinkers to cross the bridge between different levels of judgment. We are suggesting a similar process. In addition, such a pedagogy might facilitate the ability for self monitoring and control.

The importance of such analytical and remedial skills is fundamental since, as is argued in section 13.2 above, prescription of the thinking process is not appropriate or possible. It can be stated with some certainty of generalisability across all instances of clinical problem solving, that structuring and interpretation by extrapolation to stored embedding contexts is characteristic of the diagnostic thinking process. However, the occurrence of the set effect, active confirmation or any of the more specific, more idiosyncratic, more content dependent processes, is unpredictable. Examples may be constructed for demonstration, but the most useful instances might arise in the learner's own thinking processes. The teacher, therefore, should have full knowledge of the range and mechanisms of diagnostic thinking processes of neophyte and expert. From the point of view of the learner, theoretical knowledge and understanding of such cognitive processes should be helpful in facilitating self awareness and analysis and in providing a frame of reference. However, the dominant aspect of the pedagogy is neither that of adding a new subject to the curriculum nor of facilitating discussions of theory.

Rather, it is to have positive bearing upon the thinking process while it is in progress.

Clearly, a major question for the planner or designer of teaching concerns how the thinking processes of the learner might be rendered amenable to his own and the teacher's observation or analysis. A second question concerns the pedagogical means by which the range of thinking processes may be demonstrated to the learner for discussion and analysis. With regard to the first of these, the present account gathering study has demonstrated the reliability, validity and usefulness of the method of videotape stimulated recall. Such a method is amenable to modification for teaching purposes as the work of Kagan (1977) and Kagan and Krathwohl (1967) in interpersonal process recall clearly demonstrates. Ward rounds and discussions of case histories and clinical problems might also provide useful opportunity for illuminating the cognitive approach of the learner. The skill of the teacher in eliciting and analysing or facilitating self analysis of appropriate reports and accounts is crucial for such methods. As with the pedagogy of structure, the opportunity for group work is available if the pedagogical validity of vicarious experience is accepted.

However, although the proposed approaches might allow access to the learner's thinking processes in situations of varying degrees of fidelity, the question of how to integrate the results of the process into the learner's diagnostic thinking process while in operation remains unanswered. At the McMaster medical school self awareness questions are included in the clinical problems which are constructed for students to solve (Learning Resources Design Project, 1975). With problems of reasonably high fidelity such as simulated patients this method would seem likely to encourage transfer of self awareness to the real clinical problem solving context. It is possible that teaching methods may be constructed with lower fidelity problems which also may provide for transfer. It may be

necessary to determine whether or not prompts to self analysis or monitoring should best occur at fixed intervals or at critical points identified by an observer (either peer or teacher). In the light of the McMaster use of observers for simultaneous interpretations of the clinical problem solver's strategy (see section 3.1.2) this would seem a reasonable strategy. The guiding principle in all this, however, must be that self monitoring and rectification of thinking or problem solving processes can best facilitate transfer to actual clinical practice if learned in a situation which closely approximates the real one. The skill of thinking, like any other, can best be learned through a judicious combination of theory and practice (Fitts and Posner, 1967). The phases of skill learning, as defined by Fitts and Posner, (1967) commence with the learner's efforts to understand the task:

"Whether left to his own devices or tutored by an experienced instructor, the beginner in most adult skill learning situations tries to "understand" the task and what it demands. A good instructor will call his attention to important perceptual cues and response characteristics and give diagnostic knowledge of results". (p. 11)

Although Fitts and Posner (1967) are here referring to the development of complex motor skills, the analogy with the present cognitive skills development is clear. These authors also describe the learner's reactions during the phases of learning and provide a description of the development here envisaged of the diagnostic thinking process. In particular, they clarify a potential meaning of self monitoring, suggesting that this declines to a lowered, autonomous, necessary level during the final phase of learning:

"During the early phase of skill learning it is usually necessary to attend to cues, events, and responses that later go unnoticed. ... During the final phase of skill learning, component processes become increasingly autonomous, less directly subject to cognitive control, and less subject to interference from other ongoing activities or environmental distractions. In this phase, skills require less processing. This means that they can be carried on while new learning is in progress or while an individual is

engaged in other perceptual and cognitive activities". (pp 12, 14)

Thus, by analogy, it would be reasonable to infer that self monitoring and awareness can be incorporated into the diagnostic thinking process by successive stages until unobtrusively integrated. The possibility of achieving such an aim in a field of cognitive skill, rather than that of the motor skills addressed by Fitts and Posner (1967) is suggested by some other authors. For example, Lawless (1979) considers with regard to learning, that "creating awareness in the student of his learning processes and their potential for improvement is the first priority". "Saljo" (1978) agrees that "... when people become aware of their own learning in different respects, they will be better equipped to deal with various sorts of learning difficulties". Such is our thesis with regard to the development of the diagnostic thinking process.

We may now turn to the second of the two questions identified, which concerns the pedagogical means by which the range of thinking processes might be demonstrated to the learner for his intellectual understanding and subsequent analytical use. It has already been suggested that many examples will arise in the learner's own thinking processes and methods of making these amenable to observation and analysis have been suggested. However, such examples might require clarification by reference to specially prepared and constructed demonstrations which, in turn, need not be didactic but may provide the student with opportunities for guided analysis of a controlled field of information. Such demonstrations may take the form, for example, of videotapes, films, reports or accounts with designed occasions for interruption, analysis and discussion. Appropriately planned conjunctions of clinical information and the problem solver's response presented on videotape, film or by report might provide a useful controlled device for teaching and learning the elements of the diagnostic thinking process by demonstration and analysis. Demonstration of the more generalisable cognitive operations may be provided by the

learners themselves. For example, the process of extrapolation, the basis of it in selection of the forceful feature and related concepts of selective attention may be elicited by means of simple exercises such as those described in the previous section (13.3.3) whereby the learner might be presented with a field of information and asked to interpret it and explain his rationale for the way in which he does so. Conflicting interpretations may arise in a group of learners and, with the guidance of the teacher, lessons drawn about identification of the forceful feature, selective attention and so on. Such methods are not to be confused with training in hypothesis generation (see 4.2.3 above). Many media and methods are available for reinforcement, demonstration and analysis of the fundamental concepts and cognitive processes of clinical problem solving as identified throughout Chapters Eleven and Twelve and summarised at the beginning of this section. Some of the ones suggested here reflect those reported and discussed by Peel (1971) in relation to promotion of mature judgment among adolescents. We may note, finally, that less active teaching methods may sometimes be both adequate and appropriate given the necessity of some theoretical knowledge. Ausubel et al's (1978) contention that reception learning can be both meaningful and active presents the possibility of didactic exposition and economical use of teaching and learning time. Their position may be summarised:

"... reception learning is not necessarily rote in character. Much ideational material (concepts, generalisations) can be internalised and retained meaningfully without prior problem solving experience. And at no stage of development does the learner have to discover principles independently in order to be able to understand and use them meaningfully." (p. 25)

In these terms:

"... meaningful learning takes place if the learning task can be related in non-arbitrary, substantive (non-verbatim) fashion to what the learner already knows, and if the learner adopts a corresponding learning set to do so". (p. 27)

The discussion of the pedagogy of process indicates features both additional and complementary to those of the pedagogy of structure. We may therefore conclude with a brief discussion of how these two aspects of the pedagogy of the diagnostic thinking process may be related. The discussion of the psychology of the diagnostic thinking process and its development (Chapter Twelve) makes it clear that structure and process are inextricably inter-related and inter-dependent. The present discussion highlights two further aspects of the pedagogy of the diagnostic thinking process. These are its theory and its practice. It is in the practice that structure and process are integrated while in the theory they are separable. In their separate forms it is suggested that they are useful for understanding and for providing a frame of reference for analysis and for the design of teaching methods and media. Given this, they may most usefully remain separate when teaching and learning about the diagnostic thinking process. When facilitating or developing that process the two can only be integrated. Thus the same teaching methods can be used to identify and demonstrate both structure and process aspects of clinical problem solving. If the learner is undertaking interpretation of clinical information by the proposed exercises in extrapolation he is, perforce, using both process and structure. However, if we consider it more closely, the issue appears to resolve itself quite logically. Reference to the summary of the discussion of the pedagogy of structure (see the conclusion of section 13.3.3) shows quite clearly that, from the point of view of the learner, the pedagogy appears to be one of process. Structure only appears as a psychological construct in the theoretical aspects of learning which might facilitate greater self awareness and understanding. The major importance and application of our knowledge of the role of structure in the diagnostic thinking process is in relation to the design of teaching methods, the facilitation of appropriately stored learning, and as a frame of reference for the teacher in analysis and guidance of the learner's performance. Thus, the

integration of structure and process in the diagnostic thinking process is reflected in the associated pedagogy and derived teaching and learning methods. Similarly, the analytical use of these concepts is apparent in similar ways both in the study of the diagnostic thinking process and in the discussion of pedagogy and design of teaching methods facilitatory to its development.

Before proceeding to an evaluation of the proposed pedagogy, we may summarise the present discussion as follows:

1. Practice in solving clinical problems is inadequate in the absence of guidance and feedback about the cognitive processes used.
2. The pedagogy of process concerns the development of self monitoring, self awareness and cognitive correction or compensation if necessary, rather than the development of new skills. It also concerns the rational, efficient and effective development of the balance of cognitive processes typical of the experienced clinician.
3. The importance of the development of self analytical and remedial cognitive skills derives from the inappropriateness and impossibility of prescription of a diagnostic thinking process applicable in all circumstances of clinical problem solving.
4. Although some cognitive processes are generalisable across all instances of clinical problem solving, many others are more idiosyncratic, content dependent and unpredictable.
5. The teacher should have full knowledge of the range and mechanisms of the diagnostic thinking process of neophyte and expert. The learner requires sufficient theoretical knowledge to facilitate self analysis, self awareness and remedy.

6. The learner's thinking processes may be observed and analysed by means of stimulated recall and accounts elicited during discussions of case histories, ward rounds and clinical problems. The skill of the teacher is crucial in eliciting appropriate reports and facilitating self awareness and analysis. The opportunity for group work is available.
7. The skill of thinking can best be learned and transferred by a judicious combination of theory and practice.
8. The skill of self monitoring should decline to a lowered, autonomous, necessary level during the final phase of learning.
9. The range of possible thinking processes may be demonstrated by means of the methods defined in (6) above, as well as by specially constructed and designed videotapes, films, reports and accounts. These might have opportunity for interruption, analysis and discussion. Meaningful reception learning might also have a useful and appropriate role.

With regard to the inter-related and inter-dependent relationship between the pedagogy of structure and the pedagogy of process:

10. Structure and process are integrated in practice while separate in theory.
11. The same teaching methods might be used to identify and demonstrate both structure and process aspects of clinical problem solving.
12. The major role of our knowledge of structure is in the design of teaching methods and as a frame of reference for the teacher.

13.3.5 An Evaluation of the Proposed Pedagogical Approaches

An evaluation of the discussion of pedagogical approaches may be achieved in relation to the criteria discussed in section 4.1 above. These criterial areas are: the role of structure in learning; transfer of learning; and, problem solving and learning. Each of these we may discuss in turn.

For a pedagogy to be structurally sound, section 4.1.1 points out that it must enable the learner to understand the subject matter in such a way that he may relate other things to it meaningfully. The exposure which the learner has to the subject must enable him to use it in his subsequent thinking. It must admit of changes in the developing stored structures of knowledge and skill. It must facilitate an efficient and effective interaction between substance and aptitude. It must take into account the learner's existing cognitive structure. It must facilitate the increased manipulability of the body of knowledge and the generation of new propositions. It is suggested that the proposed approach pays attention to each of these criteria. Reference to the discussion of the pedagogy of structure (section 13.3.3) shows that the approach should enable the learner to appreciate the flexibility and manipulability of arrays of information. The roles and relationship of stored structures and elicited information are clarified. Information is identified in relation to different stored structures. The encouragement to identify the elements of the field and work with them by organising and reorganising them in relation to extrapolated contexts is emphasised. Conscious awareness of the role of the forceful feature and its derivation and characteristics is facilitated. Restructuring, extension or other alteration of stored structures is made easier by the process of active and analytical response to information so that relationships, connections and disjunctions are made the subject of conscious thinking. Since the extension of the store of knowledge either by accretion or some form of restructuring is based in the context of self monitoring problem solving

thinking, it is suggested that both the approach and the structures themselves will gain the necessary flexibility to be of efficient and effective use in subsequent thinking. In particular, the possibility and means of future change and development is explored and internalised as part of the learning process itself. The role, mechanism and concept of extrapolation to embedding stored structures is given its due emphasis in the proposed pedagogy such that the learner may gain the process skill itself. Indeed, section 13.3.4 makes it quite clear that process and structure are integrated in the pedagogy and separated only for analytical and design purposes. Finally, the pedagogy is not prescriptive of a process or a structure. Instead, each learner works from and takes advantage of his own existing structures through the process of conscious awareness of his own cognitive processes and their substantive outcomes. Each learner is guided towards self awareness. Thus individual differences in structural content and use may be revealed and built upon.

For the pedagogy of the diagnostic thinking process to ensure transfer of learning, section 4.1.2 points out that present learning must facilitate future problem solving. The transfer must be of structure and process and should not be speciality specific but generalisable. The learner should acquire an understanding of the fundamental structure of the subject, and a transferable self awareness, monitoring and, where necessary, rectification of cognitive error. In the context of the diagnostic thinking process, transfer of substance must also involve transfer of the capacity for restructuring stored knowledge. It is argued that the development of a conscious awareness of the cognitive operations of the diagnostic thinking process and an ability for contemporaneous self monitoring is, of necessity, transferable since some subset of the range of operations to be taught (see section 13.3.4) is characteristic of every instance of clinical problem solving. The lack of prescription mentioned earlier and the provision for individual differences in prior struct-

ures and in development of substance and aptitude also should lay a foundation, not externally or arbitrarily imposed, appropriate to each individual which is amenable to further future development. Training in the processes of diagnostic thinking and the structurally based design of substance acquisition is planned specifically to facilitate future problem solving in clinical practice. The structural characteristics of the information, for example, that of the forceful feature, are applicable, and therefore transferable to all instances of clinical problem solving. Indeed, in section 13.3.3 it is argued that undergraduate medical education may be seen as an advance organiser for clinical practice and problem solving. The active nature of the pedagogy proposed should ensure internalisation of the analytical processes and their integration with the clinical problem solving process. In section 13.3.3 it is stated that the future effects of learning are dependent upon the steps taken at the time of storage. Here, those steps concern both substance and aptitude. The learner's attention is guided towards the perception and evaluation of the items of information presented, their relationship with information stored in memory, the structural characteristics of both, their developing aspects and the cognitive processes which operate on them. In this way, the learner should acquire the necessary understanding of the fundamental structures and processes of clinical problem solving.

In terms of transfer, an important question concerns transfer across specialities. It is argued that the diagnostic thinking process differs across specialities only quantitatively, but not qualitatively (see section 12.3). The same might be said of case specificity. Thus some subset of the range of possible cognitive processes of clinical problem solving is applied in all instances and cases and some, as argued, apply to all instances and cases. The proposed pedagogy allows rehearsal or analysis of all such processes and attempts to integrate self awareness into the diagnostic thinking process

itself. The balance of theory and practice in the proposed pedagogy is appropriate to the future differences in relative contribution of the separate processes across instances of clinical problem solving regardless of speciality. In addition, it seems reasonable to suggest that the proposed pedagogy can be applied and implemented in any speciality. It is not content dependent. Although based on different reasoning, the overall positive conclusions concerning the possibility of transfer of learning in medical education is reflective of Elstein et al (1978). The type of transfer of process which we describe is analogous to that described by Bruner (1960) as "non-specific transfer or, more accurately, the transfer of principles and attitudes". Such an approach as the one here proposed which should facilitate transfer of both substance and process and the ability for further development of both, should answer the problem described by Miller et al (1961):

"But the inescapable truth is that even a course of study four times as long as the present vogue could not expose a student to every situation he might someday face. In the rapidly changing twentieth century world it is impossible to predict what is likely to happen next month, let alone next year. Thus education for transfer demands conscious and thoughtful attention".

(p. 62)

Ausubel et al (1978) describe some conditions of transfer. These we may quote at length since it is considered that the proposed pedagogy matches these conditions as well as the criteria from section 4.1.2 above:

"Transferability ... is largely a function of the relevance, meaningfulness, clarity, stability, integrativeness, and explanatory power of the originally learned subsumers ... Transfer does not take place automatically and without deliberate effort to appreciate and practise the opportunities that are present for transfer in a given learning situation. The learner must also perceive the relationship of the training to the criterial task ... Transfer can be facilitated by providing opportunity for

learning principles in as wide a variety of situations as possible by explicitly emphasising the similarity between training and criterial tasks and by presenting the latter tasks continuously or in close succession".

(pp 199 - 200)

Finally, we may evaluate the proposed pedagogy from the point of view of the relationship between problem solving and learning. Section 4.1.3 states that the product of both medical education and clinical practice is a clinical problem solver who is continuing to learn. Such a product is unavoidable since clinical problems, by their very ill defined nature (see section 12.1.1) cannot be solved by rote or reproductive means, but require productive thinking and the new application, and therefore extension, of previous learning. Thus past experience must be reorganised to fit current demands. However, even for ill defined problems, it is suggested after Peel (1967) that certain basic thought processes are involved in all. These are discussed earlier in this section. We may conclude that teaching the potentially general skills might be the most useful tactic. However, it is also noted by Peel (1967) and reflected in the current study that different subjects in varying conditions emphasise different aspects of thinking. It is therefore necessary to take into account such individual differences, to note that externally imposed forms of, for example, data collection, may have no effect upon the problem solver's actual thinking process, and to make available the range of possible processes of which the clinical problem solver may at any one time apply some subset.

It is suggested that the proposed pedagogy takes into consideration each of these aspects of the relationship between problem solving and learning. It is argued (sections 13.3.3 and 13.3.4) that such a pedagogical approach would lay a firm basis for future development and alteration, since it is itself based on the assumption that the learner represents a changing system of both substance and process. The pedagogy is aimed

at facilitating that developmental process to the greatest efficiency and effectiveness. The future, and unpredictable, problem solving performance of each individual learner is also prepared for by providing him with the substance and aptitude necessary for self awareness and monitoring of his own diagnostic thinking processes. Thus a general skill is learned with the mechanism of specific application. In this way, also, individual differences in the structure of substance, in aptitude and in their development are taken into account. The teaching methods proposed allow self analysis of genuine diagnostic thinking processes as well as separate rehearsal of specific aspects of the process. In addition, it is argued (section 13.3.4) that such analytical processes can become integrated into the dynamic clinical problem solving process itself. Therefore the learner has the opportunity to learn at both a holist and an analytical level and may rehearse the master skill as well as its sub-skills. The problem based approaches discussed in 4.2.2. above, would not seem, *prima facie*, to achieve this end. The pedagogy of structure and the pedagogy of process when combined form a pedagogy of problem solving which better match our knowledge of the diagnostic thinking processes of neophyte and expert and the developmental characteristics and experiences which transform the former into the latter.

It is sometimes illuminating to resort to a truism. Miller et al (1961) did so to effect:

"Learning how to be a doctor and being one are obviously different things". (p. 62)

It is hoped, however, that the relationship between these two states has been defined and an appropriate pedagogy indicated which will make the transformation more efficient and effective than it otherwise might be.

13.4 Summary and Conclusions

Implications of the findings of the present parallel studies are considered from the point of view of their indications for undergraduate medical education as a preparation for

clinical practice. The current relationship between these is discussed and the potential relationship considered. Four aims in the design of teaching strategies are defined. On the basis of these, possible pedagogical approaches are proposed. A practical and pragmatic approach is adopted in the light of contextual considerations, curriculum considerations, curriculum content and the role of the medical teacher. The pedagogy itself is discussed in terms of structure and process and their integration. Finally, the proposed pedagogy is evaluated in relation to the criterial areas of the role of structure in learning, transfer of learning, and the relationship between problem solving and learning.

It is concluded that the proposed pedagogy stands up to scrutiny in these lights and may be of help in the process of transformation from student to doctor.

CHAPTER FOURTEEN

Limitations of the Present Research and Indications for Further Study

The question of the limitations of the research here reported and of the indications given of a need for further study are of necessity closely related. However, for the purposes of discussion they may be dealt with separately.

14.1 Limitations of the Present Research

The limitations of the present research derive principally from two sources: firstly, the research design itself; and secondly, the variables selected for or amenable to study. The research design does not allow longitudinal sampling of the diagnostic thinking process. Therefore, conclusions concerning its development are dependent upon the validity assigned to the arguments for comparability of the samples of subjects within and between the parallel studies (see sections 5.6.1 and 5.6.3). Although those arguments are not spurious, it is undoubtedly the case that longitudinal samples would have been less open to any doubts about comparability. Likewise, the research design consists of two parallel studies of separate samples. Again, although the arguments for their comparability are not spurious, the same samples across studies would have been more elegant and satisfactory.

A second limitation of the research design concerns the lack of opportunity to study in depth individual differences within samples of subjects. For example, no analysis of the substance structure of any student, house officer or registrar is made or available. With regard to process, the account gathering study, which perforce collects information about each subject individually and one at a time, yields no descriptive or quantitative information concerning the cognitive processes of any individual subject. The method of eliciting accounts followed by content analysis yields data and results concerning counts of incidences and frequencies by group and not by individual subjects.

The third limitation of the research design concerns the comparison of specialities. It is reasonably argued and concluded on the basis of the data that speciality differences are quantitative and not qualitative, but studies of other specialities would, if they yielded results of the same order, give this conclusion a sounder base. As far as the account gathering study is concerned, no speciality based study has been made at all. Thus, the conclusions are based on logical and inferential argument and reference to the work of Barrows and the McMaster group in the field of neurology.

The final limitation of the present research concerns the range of samples selected for study. It is assumed that the samples of final year students are representative of the products of the undergraduate medical curriculum. Extension of sampling further back to second and first year clinical students, and even to second and first year pre-clinical students would give more substantial grounds for this assumption. Likewise, the present sampling ceases at the level of registrar/senior registrar. It is assumed that these subjects represent the diagnostic thinking process of the experienced practitioner, yet it cannot be assumed that they also represent the fully mature process. Such an inference or conclusion could only be based on a sample of consultant clinicians of varying clinical experience.

The second source of limitations of the present research concerns the variables selected for or amenable to study.

Although failure to study all possible variables does not invalidate reasonable conclusions, interpretations of results may be more deep and broad when further variables are added to the pool. This is so for the present study. For example, the role of the physical examination in the clinical problem solving process is not studied. Neither is the stage of patient management and treatment considered. It is argued that this does not in any way invalidate the research or conclusions based upon its results, but it is nonetheless not quite complete. A second variable not studied is the precise experience of the

subjects in undergraduate medical education. An analysis of the content and processes of teaching to which each subject or group of subjects had been exposed might have proved interesting and relevant and might have added substance to arguments concerning comparability of samples of subjects or the discussion concerning the present relationship between medical education and clinical practice (see 13.1.1). A third variable not studied is that of personality and its relationship to cognitive style or clinical problem solving approach. Simpson (1972) reports studies of psychological effects of the undergraduate medical curriculum and cites, for example, the authoritarian personality as a variable of interest. Miller et al (1961) also consider personality as a relevant variable.

A fourth variable of possible relevance but not studied concerns the possible role relatedness of the diagnostic thinking process. The tasks of student, house officer and registrar are quite different. The first has no clinical responsibilities and is very much an imbiber of knowledge and skill; the second has a task of thorough in-patient clerking and management, but not of making definitive diagnoses; while the third has greater clinical responsibility and the task of diagnosis and treatment. It is not suggested that these different roles would have great determinative effect on the quality of the diagnostic thinking processes, but rather that some variance might be found in quantitative and content aspects.

The final variable of possible direct relevance and interest which is nonetheless not studied is that of the subjects' cognitive processes in content areas other than that of clinical problem solving. It is suggested that the diagnostic thinking process is a special application of everyday cognitive operations (see Chapters Twelve and Thirteen) yet this assertion is based on general psychological theory, not on evidence of the subjects working with ill defined problems in different content areas. It would be of some substantial interpretative importance to have shown that the subjects do display the same cognitive operations in other aspects of their lives or with

other types of problem. Although other variables of possible relevance may be identified, those discussed are considered to be of the greatest importance and relevance and, therefore, have the most limiting effect upon strength and type of inferences which are drawn from the results of the parallel studies. We may now consider both these limitations and results in the light of their implications for further research.

14.2 Indications for Further Study

Each of the limitations of the research which has been discussed also constitutes an indication for further study. Additional work based on the limitations of the research design would be intended either to replicate or validate the present results or conclusions; whereas additional work based on those variables omitted from the research would be intended to deepen and broaden its conclusions. However, the results and conclusions of the research itself provide indications for further study in a number of other areas.

Firstly, and most importantly, the entire explanatory theoretical framework erected must be subject to study to determine the extent of its paramorphism. The following aspects seem to be of particular moment and salience for the validation of the model: the nature of meaningful differences between pre-diagnostic and diagnostic interpretations and their use; the nature and psychological mechanism of the selecting operation and the identification of the forceful feature; the psychological mechanism of extrapolation; characteristics or features which confer or constitute different degrees of forcefulness; cognitive reasons for and features of different degrees of satisfactoriness of extrapolation; features which confer sufficient threshold of force; variables which affect restructuring; the conditions of error; features which prompt the problem solver's determinative and responsive behaviour during the clinical interview, in particular the determinants of his precise questioning strategy; and, finally, determinants of the nature of acceptable resolution of the problem.

The results of the present study have facilitated certain inferences about appropriate pedagogy in the development of the diagnostic thinking process. Although the proposals themselves are tentative they provide the basis for further study of the most practical kind. One of the major reasons for conducting the present study is described in Chapter One as the prior necessity to understand the diagnostic thinking process and its development before being able to design appropriate teaching methods. Thus, the study is perceived primarily as enabling rational pedagogical design. It is this design which requires further definition, controlled implementation and evaluation. In addition, the question of how medical teachers might be trained and a new pedagogy introduced into existing medical curricula is of considerable importance and not without difficulty (Abercrombie et al, 1978). The present study represents a means of advance in rational pedagogical design. But much remains to be done.

14.3 Summary

The limitations of the present research are described and attributed either to the research design itself or to the selection or availability of variables for study. Such limitations also constitute indications for possible further study. Additional areas which require subsequent work are described and identified as deriving from either the psychological study or the associated inferences concerning appropriate pedagogy.

EPILOGUE

We may close with the words of a student, recorded during an account gathering session. His description reflects the experience of being a medical student and the process of learning to be a clinical doctor. In his own way, he reflects also the theoretical, structuralist rationale of this study. His words describe the process which, we hope, the present study might help rationalise and advance to good effect:

"I've got twenty minutes with this patient. When you relate these things to consultants, it's good to tell a story. Mind you, it's only interesting now, when you begin to understand something. There's a bridge, a link, you actually know something about the condition, so therefore you can look behind it".

12 December, 1976

APPENDICES

APPENDIX 1

Information Given About the Features of Syncope and Epilepsy

· Loss of consciousness may be due to syncope or major epilepsy.

Features of syncope

1. Often related to posture (upright).
2. Preceded by faintness, giddiness or nausea and sweating.
3. Patient may remember falling or be aware throughout the episode of what is happening to him, or may not.
4. Is not accompanied by violent, uncontrolled limb movements or micturition.
5. No severe after effects such as headache or prolonged sleepiness.

Features of major epilepsy

1. Not posture-related.
2. May occur following a specific 'aura' (with examples).
3. Patient will never remember falling or any of the circumstances occurring during the seizure.
4. A violent, uncontrolled and uncontrollable discharge of neuronal activity making all co-ordinated motor or cerebral activities impossible. Violent, uncontrolled jactitations, may micturate.
5. Often followed by evidence of severe cerebral disturbance such as prolonged sleepiness, automatism, paralysis or headache.

Acromegaly

- (a) Serum growth hormone measurements during a glucose tolerance test is a well recognised diagnostic procedure in acromegaly.
- (b) Acromegaly is common in children.
- (c) Patients with acromegaly may have a bitemporal hemianopia due to a pituitary tumour.
- (d) Plasma phosphate concentration is raised in acromegaly.
- (e) Headache is a symptom of acromegaly.

Hypothyroidism

- (a) Hypothyroidism in children affects bony growth and development.
- (b) A normal serum TSH level is very rare in untreated myxoedema.
- (c) Hypothyroidism occurs equally in men and women.
- (d) Localised pre tibial myxoedema is common.
- (e) There is an increased incidence of pernicious anaemia in patients with hypothyroidism.

Primary Hyperparathyroidism

- (a) In hyperparathyroidism the serum calcium and phosphorous are elevated.
- (b) Diarrhoea due to irritant colitis is common.
- (c) Many newly diagnosed cases nowadays are asymptomatic.
- (d) The disease causes skeletal deformities in some patients.
- (e) Hyperparathyroidism is due to a malignant tumour of the parathyroid in the majority of cases.

Seminiferous Tubule Dysgenesis (Klinefelter Syndrome)

- (a) The patient may complain of gynaecomastia coming on at puberty.
- (b) On examination the testes are characteristically normal.
- (c) Patients are usually short.
- (d) These patients are characteristically infertile.
- (e) All of the cells of these patients contain two X chromosomes in most cases.

Cushings Syndrome

- (a) Patients with endogenous Cushings syndrome can have bilateral adrenal hyperplasia.
- (b) The pituitary fossa is enlarged in pituitary dependent Cushings syndrome in the vast majority of cases.
- (c) Patients with untreated Cushings syndrome which is not due to a carcinoma very rarely die from the disease in a five year period.
- (d) Proximal muscle weakness is rare in patients with Cushings syndrome.
- (e) Patients with Cushings syndrome due to adrenal adenoma usually suppress the plasma cortisol if given 8 mg dexamethasone daily.

Auto Immune Thyroiditis

- (a) In auto immune thyroiditis there is a positive correlation between lymphocytic infiltration of the thyroid gland and thyroid auto antibody titres.
- (b) Auto immune thyroiditis can be familial.
- (c) Auto immune thyroiditis can be the cause of a painful and tender thyroid gland.
- (d) Absence of a goitre rules out the diagnosis of auto immune thyroiditis.
- (e) The tanned red cell test does not detect antibodies to thyroglobulin.

Addisons Disease

- (a) Some patients with Addisons disease have circulating adrenal auto-antibodies.
- (b) Basal plasma cortisol concentration at 9 am may be 6 µg% or greater.
- (c) Pigmentation of the skin may occur.
- (d) Postural hypotension occurs.
- (e) In Addisons disease the excretion of water is impaired.

Ovarian Dysgenesis (Turner's Syndrome)

- (a) The patients are characteristically infertile.
- (b) Cardiovascular abnormalities occur in some cases.
- (c) The majority of these females are chromatin positive.
- (d) Retarded bone age leads to excess height.
- (e) These patients usually present with secondary amenorrhoea.

Phaeochromocytoma

- (a) Phaeochromocytomas are sometimes bilateral and malignant.
- (b) Persistent hypertension can occur with phaeochromocytoma.
- (c) Profound hypotension may follow surgical removal of previously undiagnosed cases.
- (d) Urinary excretion of 5HIAA is a useful screening test.
- (e) There is an association with medullary carcinoma of the thyroid in some patients.

Section B

Hypoglycaemia

A patient is said to be in hypoglycaemic coma from an overdose of soluble insulin. Which of the following findings are compatible with this?

- (a) Gradual onset of unconsciousness over 1-2 days.
- (b) Clinically evident dehydration.
- (c) Acidotic breathing pattern ("air hunger").
- (d) Dilated pupils but normally reactive pupillary reflexes.
- (e) A pulse rate of 120/minute and pale sweaty skin.

Hypercalcaemia

A patient has a serum calcium of 17 mg%. The following laboratory and clinical findings may be diagnostically related to the hypercalcaemia.

- (a) Bilateral hilar lymphadenopathy on chest x ray.
- (b) Hb S found on haemoglobin electrophoresis.
- (c) A history of coeliac disease.
- (d) A mastectomy two years previously for carcinoma of the breast.
- (e) An abnormal 'M' band on serum protein electrophoresis.

Diabetes Insipidus

In diabetes insipidus the following may be consistent with the diagnosis:

- (a) A body weight loss of greater than 5% occurs when the patient is asked to stop drinking for 12 hours.
- (b) The patient complains of nocturia.
- (c) Clinical hypogonadism is found in association.
- (d) The condition came on after a head injury.
- (e) The basal plasma osmolality is found to be low before treatment with ADH is given.

Infertility

In an infertile man or woman the following are expected or characteristic findings.

- (a) Vaginal bleeding occurred after exogenous progestogen administered to a patient with Turner's syndrome.
- (b) A male patient with a chromosome complement of 47 XXY has a sperm count of 80 million/ml with 15% abnormal forms.
- (c) An enlarged pituitary fossa was found in an infertile female who complained of persistent breast discharge.
- (d) The occurrence of ovulation in a female patient was confirmed by measuring the plasma oestradiol concentration.
- (e) The occurrence of ovulation was satisfactorily confirmed in one patient when records shows she menstruated at four to six week intervals.

Short Stature

A boy of 15 is referred with short stature. It may be true that:

- (a) The height of the boy's parents can be used as one means of estimating his height expectancy.
- (b) A history of chronic otitis media indicates a pituitary cause.
- (c) If his height is on or above the third percentile an abnormality of growth is excluded.
- (d) If he has no evidence of puberty this suggests that his bone age will be less than 14 years.
- (e) A history of diarrhoea from early childhood can be diagnostically related to his short stature.

Gynaecomastia

In gynaecomastia the following findings can readily be related by indicating an underlying diagnosis or cause.

- (a) A history of heavy drinking.
- (b) A history of haemoptysis and weight loss.
- (c) A varicocele on the right hand side.
- (d) High arched palate and long fingers.
- (e) Both testes are 1 cm. in diameter but firm.

Delayed Puberty

An otherwise normal boy of nearly 16 is referred because his doctor considers his puberty is delayed on the grounds that his voice has not broken and he has a feminine appearance. It is entirely consistent with this clinical picture that:

- (a) He has a bone age of 16 years with no clinical evidence of puberty on examination.
- (b) The boy has no beard development but scrotal rugosity and some pubic hair are present.
- (c) His chromosome complement is 45 X O.

- (d) His span is 6 feet and his height is 5'9" and the testes are 1 cm. in diameter.
- (e) The boy is found to have no sense of smell.

Goitre

An English school girl of 18 presents with a large goitre.

- (a) Dietary iodine deficiency is a likely cause.
- (b) A finding of deafness may be related to the diagnosis.
- (c) Hypothyroidism may be present.
- (d) A family history of goitre may be related to the diagnosis.
- (e) Recent weight loss of twenty pounds may be diagnostically related to the goitre.

Hypopituitarism

A man of 18 is referred by his optician following the finding of left optic atrophy. He is 5'2" tall with no beard and a dry pale skin. The testes are small. It is consistent with this clinical picture that:

- (a) Clinical dehydration is present.
- (b) A low plasma aldosterone concentration is found.
- (c) An elevated serum TSH concentration is found on investigation.
- (d) He has a bone age of approximately 18 years.
- (e) He needs a larger than normal dose of insulin to induce hypoglycaemia in an insulin stress test.

Tetany

A 16 year old school girl presents in casualty with flexor spasms of both wrists after a row with her boyfriend. The thumbs are adducted.

- (a) A history of anaemia and offensive diarrhoea may be diagnostically relevant.
- (b) A disturbance of respiration may be involved in the diagnosis.
- (c) Previous surgery to the neck should be excluded.

- (d) The finding of a positive Trousseau's sign is entirely compatible with the clinical picture.
- (e) The serum calcium concentration may be normal.

Virilism

A 24 year old waitress is concerned by the growth of hair on the forearms, breasts and thighs for 18 months. She has secondary amenorrhoea of 2 years duration, some deepening of the voice and enlargement of the clitoris.

- (a) The information given is entirely consistent with a diagnosis of constitutional hirsutism.
- (b) The information given is consistent with a diagnosis of adrenal tumour.
- (c) The information given is consistent with an ovarian tumour.
- (d) She may correctly be described as virilised on the information given so far.
- (e) A high urinary pregnanetriol excretion would suggest an ovarian tumour.

Obesity

In these patients:

- (a) A calorific intake in excess of expenditure occurs.
- (b) Carbohydrate restriction should be advocated for 3 days before assessing glucose tolerance in the obese state.
- (c) Cushing's syndrome should be considered if the arms and legs are thin in respect to the trunk.
- (d) You would expect the incidence of diabetes mellitus to be greater than in non obese adults.
- (e) A restrictive type of result in pulmonary function tests occurs in severe cases.

Section C

Case A

A 57 year old widow who lived alone consulted her doctor in July because of fatigue. On examination she looked pale and he prescribed ferrous sulphate 200 mg tds. In 3 weeks she returned complaining that she felt no better and in fact the tablets had greatly exacerbated her constipation so that she had to take Epsom Salts. Her doctor requested a blood count which showed Hb 10.4 G/dl. White cell count 6400/cu MCHC 32, MCV 110 and platelets 240,000/cu. He referred her to the clinical haematologist at the local hospital where a sternal marrow showed normoblastic erythropoiesis and a biochemical screen of serum revealed a blood urea of 45 mg/dl, serum albumen of 4.2 G/dl and a serum cholesterol of 400 mg%. The haematologist referred the patient to a general physician who discovered vitiligo on the arms and legs. Her eyes were watering and there was extensive periobital and minimal ankle oedema. The jugular venous pressure was markedly raised though the lung bases were clear. The heart sounds were extremely difficult to hear and a chest x ray revealed marked enlargement of the heart. Urine examination showed a trace of proteinuria. He arranged for the patient to be admitted to hospital but she did not turn up and when her own doctor went round to her home to find out why, she seemed indifferent to his chiding remaining sitting by the fire when he entered the room and not rising when he left.

There is evidence of the following in this patient:

- (a) vitamin B12 deficiency
- (b) hyper magnesemia
- (c) auto immune disorder
- (d) nephrotic syndrome
- (e) pericardial effusion

Which of the following could explain this patient's condition?

- (a) carcinoma of bronchus

- (b) depressive illness
- (c) hypothyroidism
- (d) cardiomyopathy
- (e) ischaemic heart disease.

Which of the following investigations may be helpful in this case?

- (a) cardiac ultra sound
- (b) thyroid stimulating hormone assay
- (c) electroencephalogram
- (d) IVP
- (e) serum magnesium estimation.

Section D

What is the one most likely diagnosis?

Section C

Case B

A 3 year old girl following a cold and sore throat suddenly was noticed to become very thirsty. She also began to wet the bed and had to pass water several times a night. In the past few weeks, although her appetite was normal, she had lost weight. She was also becoming very tired and fretful. In the past few days she had a productive cough. Bowels were normal.

On examination the child was fretful and looked tired. Her skin was dry and there had been some loss of weight. No anaemia nor lymphadenopathy. The pulse was 92 per minute regular, blood pressure 95/60. All peripheral pulses were present. There were scattered rhonchi over the lower left chest. Respiratory rate was 24 per minute. There was some tenderness over the right iliac fossa but no guarding. The central nervous system was normal on examination. Bones and joints were normal.

There is evidence of the following in this patient:

- (a) urinary tract infection
- (b) impairment of adrenal function
- (c) hysterical high water intake
- (d) acute bronchitis
- (e) high secretion of thyroid hormone.

Which of the following could explain this patient's condition?

- (a) thyrotoxicosis
- (b) Addisons disease
- (c) acute peritonitis
- (d) diabetes mellitus
- (e) urinary tract infection.

Which of the following investigations may be helpful in this case?

- (a) intravenous pyelogram
- (b) serum triiodothyronine concentration
- (c) urine test for glucose
- (d) plasma bicarbonate concentration
- (e) chest x ray examination.

Section D

What is the one most likely diagnosis?

Section C

Case C

A 17 year old with slight mental subnormality was brought to the clinic by his father, because of failure to "thicken out". The boy worked as a gardener for the local council. Appetite and food intake were normal and the bowel habit a little constipated. There were no other symptoms except occasional headaches but the father had once had pulmonary tuberculosis. The mother had died aged 42 from a heart

complaint. The patient was on the 25th percentile for height but the 3rd percentile for weight. On examination there was no beard, leg or chest hair but a little pubic and some axillary hair were present and there was minimal gynaecomastia. The testes were 1 cm. in length. A soft mid systolic murmur could be heard which varied with posture but there were no other abnormalities in the cardiovascular system. Both arms and legs however seemed rather long for his height. Mobility was normal.

There is evidence of the following in this patient:

- (a) a chromosome abnormality
- (b) eunuchoid habitus
- (c) giantism
- (d) congenital heart disease with somatic retardation
- (e) hypogonadism.

Which of the following could explain this patient's condition?

- (a) testicular ectopia with maldescent
- (b) Klinefelters syndrome
- (c) congenital syphilis
- (d) Marfans syndrome
- (e) mumps.

Which of the following investigations may be helpful in this case?

- (a) serum FSH concentration
- (b) examination of chromosome morphology
- (c) serum testosterone concentration
- (d) somatomedin assay
- (e) bone age estimation.

Section D

What is the one most likely diagnosis?

Section C

Case D

A man of 52, a heavy smoking and drinking floor tile layer, attended the clinic with 1 stone weight loss over the last 3 months and tender painful wrists and ankles. He had had an irritating cough for 6 months and in the last 3 weeks had had to stop drinking his usual 3-4 pints of beer at night because he felt unwell after the first pint. He had also been somewhat forgetful of late. On examination there was clubbing of the fingers and some oedema of the ankles. Examination of the chest showed a kyphosis of the thoracic spine. The liver edge was just palpable and pulsation was present in the epigastrium. He was admitted to the ward where the next day he suffered an epileptic fit. A brain scan was normal but a chest x ray showed a rounded but uncalcified opacity in the left mid zone 2 cm. in diameter. The serum sodium was found to be 105 mEq/l.

There is evidence of the following in this patient:

- (a) metastatic cancer
- (b) hypercalcaemia
- (c) ectopic hormone formation
- (d) negative water balance
- (e) hepatocellular failure.

Which of the following could explain this patient's condition?

- (a) bronchial carcinoma without metastases
- (b) rheumatoid arthritis
- (c) inappropriate ADH secretion
- (d) alcoholic hypoglycaemia
- (e) hepatic cirrhosis.

Which of the following investigations may be helpful in this case?

- (a) LE cell preparation
- (b) bilateral carotid arteriography

- (c) simultaneous urine and serum osmolality estimations
- (d) bronchoscopy
- (e) glucose tolerance test.

Section D

What is the one most likely diagnosis?

Section C

Case E

A 62 year old lady was admitted with a history of haematuria on three occasions in the last month. Each episode had been accompanied by right loin pain. She had also noticed during the past year aching pains in her muscles and back. She also suffered from abdominal pain after meals and on opening her bowels and nocturia four times per night. Physical examination showed an agitated lady with a red face. Pulse was 100 per minute, BP. 170/105. Cardiovascular system was otherwise normal. Examination of the back showed several papillomata. Abdominal examination revealed a loaded sigmoid colon. Cranial nerves were normal, all reflexes were present and brisk and the plantar responses were flexor.

There is evidence of the following in this patient:

- (a) hypercalcaemia
- (b) malignant carcinoid syndrome
- (c) severe endogenous depression
- (d) renal calculi formation
- (e) polymyalgia rheumatica.

Which of the following could explain this patient's condition?

- (a) osteomalacia
- (b) myxoedema
- (c) osteoporosis
- (d) primary hyperparathyroidism
- (e) bladder papilloma.

Which of the following investigations may be helpful in this case?

- (a) x ray of the hand
- (b) IVP
- (c) plasma alkaline phosphatase
- (d) blood urea estimation
- (e) water deprivation test.

Section D

What is the one most likely diagnosis?

* Copyrighted

APPENDIX 2

Questionnaire in Endocrinology - Content*

Section A

In Juvenile Onset Diabetes Mellitus

- (a) Endogenous insulin secretion is frequently normal.
- (b) The condition may come on acutely, over a few days.
- (c) Severe dehydration occurs in some untreated patients.
- (d) Lack of insulin precipitates the development of ketosis.
- (e) Visual disturbance due to refractive changes occurs as a symptom.

Hyperthyroidism

- (a) In Graves' disease true exophthalmos is only found in a minority of cases.
- (b) In Graves' disease long acting thyroid stimulator-protector (LATSP) may be found in the blood stream.
- (c) Atrial fibrillation without other clinical evidence of hyperthyroidism may occur in the elderly.
- (d) Graves' disease is commoner in females.
- (e) It is rare for proximal muscle weakness to be associated with hyperactive reflexes in Graves' disease.

Diabetes Mellitus (Maturity Onset)

- (a) Patients with maturity onset diabetes mellitus commonly present without ketosis.
- (b) Patients may respond to treatment with diet alone.
- (c) Treatment with steroidal drugs may precipitate the onset of diabetes.
- (d) Neuropathy does not normally occur in patients with maturity onset diabetes.
- (e) The kidneys are not usually affected.

APPENDIX 3

Questionnaire in Neurology - Content*

1. IN VITAMIN B12 DEFICIENCY

- (a) Neurological complications only occur when the haemoglobin falls below 10 gm%.
- (b) Folic acid improves the neurological symptoms but does not help the anaemia.
- (c) The lateral spinothalamic tracts are not involved.
- (d) The plantars may be upgoing even when the knee jerks and ankle jerks are depressed.
- (e) Dementia or optic atrophy occur in a minority of cases.

2. REGARDING PARKINSONISM

- (a) In Parkinson's disease there is depigmentation of the red nucleus.
- (b) Rigidity is of the spastic type.
- (c) Symptomatic Parkinsonism is most commonly due to phenothiazines.
- (d) L-dopa crosses the blood-brain barrier and is converted to dopamine, the active pharmacological agent.
- (e) The tremor is present at rest and is decreased by voluntary movements.

3. IN MIGRAINE

- (a) The headache is always over one superficial temporal artery.
- (b) Attacks may be related to the menstrual cycle.

- (c) Visual symptoms may include flashing lights, hemianopia and scotomata.
- (d) Attacks usually last less than one hour.
- (e) Permanent neurological sequelae do not occur.

4. REGARDING MULTIPLE SCLEROSIS

- (a) It occurs more commonly in the tropics.
- (b) In the middle-aged may present as a slowly progressive paraparesis without relapses and remissions.
- (c) Rarely affects the optic nerves.
- (d) There is a mean survival greater than ten years.
- (e) The findings of an abnormal Lange curve, but negative Wassermann Reaction is in favour of the diagnosis.

5. REGARDING INTRACRANIAL HAEMORRHAGE

- (a) Chronic subdural haematoma is apt to occur in the elderly and in alcoholics.
- (b) In extradural haemorrhage a dilating pupil suggests tentorial herniation.
- (c) Intracerebral haemorrhage usually has a stuttering onset.
- (d) Surgery for cerebral aneurysm should be delayed for at least six weeks after a bleed.
- (e) Patients with arteriovenous malformations may present with migraine or epilepsy rather than a bleed.

6. REGARDING CEREBRAL TUMOURS

- (a) The commonest type is a meningioma.
- (b) They usually have a rapid onset (less than 24 hours).

- (c) There may be no focal symptoms, only those due to raised intracranial pressure.
- (d) May be clinically indistinguishable from a cerebral abscess.
- (e) Posterior fossa tumours are more common in children.

7. REGARDING MENINGITIS

- (a) The CSF sugar is usually lowered in viral meningitis.
- (b) The three commonest causes of bacterial meningitis are pneumococcus, meningococcus and haemophilus influenzae.
- (c) E coli meningitis usually occurs in the neonate.
- (d) Can present as pyrexia of unknown origin in children.
- (e) Steroid therapy and lymphomas predispose to tuberculous and fungal meningitis.

8. REGARDING LESIONS DUE TO COMPRESSION BY SKELETAL STRUCTURES

- (a) The pain and tingling of carpal tunnel syndrome often comes on at night.
- (b) Cervical ribs may cause wasting of the small muscles of the hand.
- (c) The commonest site of lumbar disc protrusions is L 2-3.
- (d) Lumbar disc protrusions may present as acute retention.
- (e) There may be quite severe cervical spondylosis on x-ray without symptoms.

9. REGARDING EPILEPSY

- (a) A normal EEG excludes the diagnosis.
- (b) "Absences" occurring for the first time in an adult are usually not due to petit mal.

- (c) Seizures may occur during sleep.
- (d) Attacks starting in one part of the body suggest a focal lesion.
- (e) A child with one epileptic parent has a 25% chance of developing epilepsy.

10. REGARDING MYASTHENIA GRAVIS

- (a) Ptosis and diplopia are common presenting features.
- (b) Fatiguability and fluctuation in severity are important diagnostic features.
- (c) It may be associated with thyrotoxicosis.
- (d) Intravenous injection of a short acting anticholinesterase is a useful diagnostic test.
- (e) May be associated with a tumour of the thymus gland.

11. REGARDING THE COMPLICATIONS OF ALCOHOLISM

- (a) In Wernicke's encephalopathy 6th nerve palsies are frequently seen.
- (b) Korsakoff's psychosis consists of episodic paranoia and delusions of grandeur.
- (c) Alcoholism is a common cause of peripheral neuropathy.
- (d) Delirium tremens is a benign self limiting syndrome.
- (e) Tuberculosis is more common amongst alcoholics.

12. REGARDING DRUG THERAPY

- (a) Epanutin (Phenytoin) is the drug of choice for petit mal.
- (b) Tegretol (Carbamazine) helps the majority of patients with trigeminal neuralgia.
- (c) The spectrum of ampicillin includes the commonest bacteria producing meningitis.

- (d) Dementia or a psychiatric history are contraindications of the use of L-dopa.
- (e) Methysergide (Deseril) is effective for an acute attack of migraine.

Section B

1. REGARDING UPPER AND LOWER MOTOR NEURON LESIONS

- (a) In lower motor neurone lesions the tone is decreased and the reflexes are depressed.
- (b) In upper motor neurone lesions wasting and fasciculation may be seen.
- (c) In lower motor neurone lesions of the face the upper part is usually spared.
- (d) The abdominal reflexes are usually preserved in upper motor neurone lesions.
- (e) An upgoing plantar may be the only sign of an upper motor neurone lesion.

2. IN CEREBELLAR LESIONS

- (a) Inco-ordination occurs on the opposite side to the lesion.
- (b) In middle lesions there may be no nystagmus and a normal finger-nose test.
- (c) Tone is usually increased.
- (d) The tremor is made worse when a movement is performed.
- (e) Inco-ordination is present whether the eyes are open or closed.

3. REGARDING EXTRAPYRAMIDAL DISORDERS

- (a) There are marked difficulties with voluntary movements but usually no actual muscle weakness.
- (b) Sensory abnormalities may be marked.

- (c) Increased tone in association with normal reflexes may be seen.
- (d) Typically an intention tremor occurs.
- (e) In chorea quasi-purposive involuntary movements are seen.

4. REGARDING THE SENSORY SYSTEM

- (a) Position and vibration fibres cross to the opposite side, shortly after entering the spinal cord, pain and temperature fibres do not cross till the brain stem.
- (b) In peripheral neuropathy the sensory loss is most marked distally.
- (c) Loss of two point discrimination and astereognosis may be the only abnormalities in cortical lesions.
- (d) Root pain may be referred to the muscles supplied by that root.
- (e) Ataxia may be due to sensory loss.

5. REGARDING CRANIAL NERVE LESIONS

- (a) In a IIIrd nerve lesion the pupil does not react to light but reacts to accommodation.
- (b) In cavernous sinus lesions the cranial nerves involved may be I, II, III, IV, V, VI.
- (c) In cerebello-pontine angle lesions the cranial nerves involved may be V, VI, VII, VIII.
- (d) Lesions of the superior and inferior oblique are most evident when the affected eye is tested in the abducted position.
- (e) Pituitary tumours typically produce a binasal field defect.

6. REGARDING THE SPINAL NERVES AND ROOTS

- (a) A T.I. lesion produces wasting of all the small muscles of the hand but no sensory loss in the hand.

- (b) A lesion of the deep branch of the ulnar nerve causes wasting of the interossei but spares the thenar and hypothenar eminences.
- (c) A lesion of the deep palmar branch of the median nerve produces sensory loss in the 4th and 5th fingers.
- (d) An S.I. lesion produces foot drop.
- (e) A lesion of the lateral popliteal nerve produces depression of the ankle jerk.

7. REGARDING CEREBRAL LESIONS

- (a) Unilateral anosmia is suggestive of a temporal lobe lesion.
- (b) In cortical blindness the pupils are usually normal.
- (c) Parasagittal lesions can cause weakness and sensory loss in both legs.
- (d) The association of hemianopia, hemianaesthesia and hemiplegia suggests a lesion of the internal capsule.
- (e) The presence of a grasp reflex in an adult suggests a lesion of the frontal lobe on the opposite side.

8. IN THE COMATOSE PATIENT

- (a) The most reliable guide to progress is the briskness of the tendon jerks.
- (b) If there are focal signs the cause cannot be metabolic.
- (c) A pupil that starts to dilate is suggestive of tentorial herniation.
- (d) Overdose is the commonest cause seen in a medical ward.
- (e) High blood pressure may be due to intracranial lesions.

9. IN CASES OF RAISED INTRACRANIAL PRESSURE

- (a) Headache is worse in the erect posture.
- (b) 6th nerve lesions can occur as false localising signs.
- (c) Papilloedema occurs early in posterior fossa lesions.
- (d) Papilloedema is associated with early visual loss.
- (e) If there are no haemorrhages in the fundi lumbar puncture can be safely undertaken.

10. REGARDING THE HISTORY IN NEUROLOGICAL PATIENTS

- (a) In patients with recurrent transient neurological disturbance epilepsy, migraine and transient ischaemic attacks are common causes.
- (b) In dementia ability to remember distant events is usually affected less than ability to remember recent events.
- (c) Onset of symptoms within minutes is suggestive of infection.
- (d) Neurosyphilis can be excluded if the patient's attack of venereal disease was treated with penicillin.
- (e) In trigeminal neuralgia the pain is constant, throbbing and may last several hours.

Section C

CASE A

A 60 year old labourer was admitted following a series of fits. He was a heavy smoker and drinker. His wife reported that while they were watching T.V. one side of his body began to twitch (she thought the right side). Then he had violent movements of all four limbs, went blue and frothed at the mouth. The movements ceased after about a minute, but he did not wake up for a further fifteen minutes. He regained consciousness as the ambulance arrived but then had two further fits on the way to hospital. His wife stated that he had been generally unwell for about a month. He had been very tired, complaining of headaches and had had a nasty cough with rusty sputum.

On questioning she also remembered that he had been waking with night sweats and she thought he had lost several pounds in weight during this period.

By the time of admission to the ward his fits had ceased. The patient was thin, unkempt, looking older than his 60 years. His temperature was 38.5°C . He was very restless and irritable. He occasionally groaned and complained of pain in his head, but no coherent history could be obtained from him. Slight cyanosis was noted and on examination of his lungs there was evidence of consolidation of the right lower lobe, as well as some generalised rhonchi. His blood pressure was 170/90, but there were no other abnormalities in the cardiovascular system. The liver was enlarged to 3 finger-breadths below the right costal margin. There were no abnormalities in the cranial nerves except that fundal examination showed blurring of the disc margins with venous engorgement. He did not move the right side of the body as freely as the left side, the reflexes were increased on the right with an upgoing plantar reflex. Sensation to pain also appeared to be diminished on the right. Co-ordination appeared normal.

1. The following could be said concerning this patient :

- (a) he has grand mal epilepsy without evidence of focal content.
- (b) he has evidence of raised intracranial pressure.
- (c) he has evidence of acute pulmonary embolism in the lungs.
- (d) clinical findings suggest brainstem involvement.
- (e) clinical findings suggest a lesion of the left hemisphere.

2. The following conditions could produce the neurological abnormalities of this patient :

- (a) hypertensive encephalopathy.
- (b) cerebral abscess.
- (c) cerebral metastasis.

- (d) hepatic precoma.
- (e) tuberculosis

3. The following investigations should be performed :

- (a) lumbar puncture.
- (b) brain scan
- (c) E.E.G.
- (d) lung scan.
- (e) liver scan.

Section D

4. For this case write down the following :

- (a) site of lesion.
- (b) side of lesion.
- (c) disease.
- (d) primary site.

Section C

CASE B

A 50 year old hypertensive taxi driver was admitted to hospital. Following an argument with a passenger he was telling a friend about it over a cup of tea when he suddenly complained of severe pain at the back of his head, began vomiting and shortly afterwards lost consciousness. His wife brought in his tablets. He had been on methyl dopa 250 mgm tds. for two years. He said he had had a sore throat and been a little off colour for a few days.

On examination he was drowsy and confused. There was limitation of neck flexion and Kernig's sign was positive. He could not open the right eye. On raising the lid the right eye was seen to be turned outwards. He could not turn his right eye upwards or inwards. Light shone into either eye produced only contraction of the left pupil which was smaller than the right. The fundi showed arterial tortuosity and A-V nipping but no haemorrhages or exudates. There

were no other neurological abnormalities. On general examination his temperature was 38.0°C and his blood pressure was 220/140. His throat was reddened but there were no other abnormalities.

1. There is evidence of the following in this patient :

- (a) meningeal irritation.
- (b) Horner's syndrome.
- (c) internuclear ophthalmoplegia.
- (d) third nerve palsy.
- (e) raised intracranial pressure.

2. The following could explain this man's clinical picture :

- (a) Migraine.
- (b) Hypertensive encephalopathy.
- (c) Subarachnoid haemorrhage
- (d) syncope.
- (e) meningitis.

3. The following investigations should be performed :

- (a) fluorescein angiography of the fundus.
- (b) lumbar puncture.
- (c) throat swab and culture.
- (d) air encephalogram.
- (e) carotid angiogram.

Section D

4. For this case write down the following :

- (a) site of lesion.
- (b) side of lesion.
- (c) disease.
- (d) underlying disease

Section C

CASE C

A 36 year old housewife was admitted for investigation of weakness in the legs. She had recently been in good health but two years previously had seen her doctor about numbness and pins and needles in the right arm and leg which had cleared up without any specific treatment after about six weeks. Over the last four weeks she had noticed that her legs had become progressively stiffer and weaker and that she would tend to fall if she stumbled in the street. Over the last two weeks she had also complained of numbness in both legs. In the four days before admission she complained of symptoms of urgency and frequency.

On examination there was marked weakness and stiffness of both legs. The reflexes in the legs were increased and both plantars were upgoing. Sensation to pin prick was diminished below the level of the umbilicus. The rest of the sensory examination was considered normal, as was examination of the upper limbs and cranial nerves. General examination revealed no further abnormality except that the bladder was palpable above the pubic symphysis.

1. There is evidence of the following in this patient :

- (a) spastic paraparesis.
- (b) "stocking" anaesthesia.
- (c) thoracic sensory level.
- (d) neurogenic bladder disturbance.
- (e) motor root lesions.

2. The following could explain this patient's picture :

- (a) subacute polyneuropathy.
- (b) multiple sclerosis.
- (c) invasion of the lumbosacral plexus by cancer.
- (d) tumour compressing the spinal cord.
- (e) tabes dorsalis.

3. The following investigations should be performed :

- (a) E.M.G.
- (b) plain x-rays of thoracic spine.
- (c) myelogram.
- (d) pelvic examination under anaesthesia.
- (e) vertebral angiogram.

Section D

4. For this case write down the following :

- (a) site of lesion.
- (b) disease.

Section C

CASE D

A 61 year old man consulted his doctor because over the past two months he had noticed weakness and thinning of the muscles of the left hand. He had not experienced pain or numbness in the hand. Apart from his hand he felt well.

On examination there was wasting and weakness of all the small muscles of the left hand. The other muscles in the upper limbs appeared normal. The tone and reflexes were also normal. Some fasciculation was observed in the left thenar eminence. Examination of the legs showed normal power and tone but the reflexes were very brisk and the plantars upgoing. No sensory abnormalities were detected.

1. There is evidence of the following in this patient :

- (a) upper motor neurone lesion in the legs.
- (b) lower motor neurone lesion in the arm.
- (c) wasting in muscles supplied by both medial and ulnar nerves.

- (d) posterior column loss.
- (e) wasting in the T.1 distribution.

2. The following could explain this patient's picture :

- (a) carpal tunnel syndrome.
- (b) brachial neuritis.
- (c) cervical spondylosis.
- (d) motor neurone disease.
- (e) multiple sclerosis.

3. The following investigations should be performed:

- (a) E.M.G.
- (b) x-rays of the cervical spine.
- (c) myelogram.
- (d) carotid angiogram.
- (e) brain scan.

Section D

4. For this case write down the following :

- (a) site of lesion.
- (b) side of lesion.
- (c) disease.

Section C

CASE E

A 38 year old man presented to his doctor because of occipital headaches. For the past year he had complained of ringing noises in the right ear and poor hearing in it. He had had his ears syringed once to remove wax but with only a little improvement in hearing. He also complained of a continual vague feeling of unsteadiness. He had previously been in good health.

On examination he was found to have very poor hearing in the right ear. The tuning fork was found to be better heard near the right ear than on the right mastoid. A tuning fork placed at the vertex was better heard in the left ear. Both on smiling and raising the eyebrows there was weakness of the right side of the face. Stimulating the right cornea did not produce a blink in either eye. Horizontal nystagmus was present. The rest of the neurological examination revealed some inco-ordination of the right arm and leg. The general medical examination was normal. Quite a lot of wax was seen in both ears.

1. There is evidence of the following in this patient :

- (a) nerve conduction deafness on the right side.
- (b) upper motor neurone facial weakness.
- (c) involvement of 5th nerve.
- (d) left temporal lobe involvement.
- (e) cerebellar involvement on the right side.

2. The following could explain this patient's picture :

- (a) impacted wax.
- (b) Menière's disease.
- (c) acoustic neuroma.
- (d) basilar artery aneurysm.
- (e) Bell's palsy.

3. The following investigations should be performed :

- (a) audiometry.
- (b) E.E.G.
- (c) x-rays of the internal auditory meati.
- (d) air encephalogram.
- (e) vertebral angiogram.

Section D

4. For this case write down the following :

(a) site of lesion.

(b) side of lesion.

(c) disease.

* Copyrighted

APPENDIX 4

Rationale for Selection of Five Option Independent True False Multiple Choice Questions

Multiple choice questions may take a number of different forms but, in essence, consist of an item stem, which may be a statement or a question, and two or more responses, or possible answers to the item stem; these we shall call 'options'. Options which the candidate should mark 'correct' are called 'answers', while remaining responses are 'distractors'. Depending upon the format of the item, the candidate may be required only to indicate correct options. This type of item most commonly takes the 'one from five' format, where the candidate must select only one option as the answer. Alternatively, the candidate may be required to mark all options as either true or false. Much less straightforward formats which require the candidate to perform tasks more complex than indicating either true or false are also available (see Fleming et al, 1976) but were rejected here as requiring a degree of test sophistication unlikely to be found in British students and registrars.

Finally selected was the format in which each item has a stem and five options, each of which the candidate may mark as either true or false in relation to the stem. This format is known by different names; for example, 'multiple true false' (Hill and Woods, 1974), 'indeterminate' (Wood, 1977), 'multi-facet' (Schools' Council, 1965), 'independent true false' (Fleming et al, 1976). Such items should be, in essence, no different from a series of separate true/false questions (Harden et al, 1976; Jolly, 1976). This format was selected for a number of reasons.

Firstly, the format is extremely flexible and versatile. As Fleming et al (1976) point out, each option requires to be identified as true or false on its own merits, rather than having one option identified as correct while the others are relegated to the status of distractor by default. Thus each item requires as many decisions from the candidate as there are options.

Secondly, the format requires very little test sophistication on the part of the candidate.

Thirdly, the conventions and limitations of item writing for this format allowed sufficient latitude to test our subjects' knowledge and thinking skills adequately and appropriately in each of the four sections of the questionnaires.

Finally, this format has been extensively used in British undergraduate and postgraduate medical education (Harden et al, 1976) and so should have been familiar to most of our subjects. In particular, our subjects were largely drawn from the medical schools and hospitals of London University (where this format is used) and Newcastle (where the format also originated). In addition, the format has been used by The Royal College of Physicians, of which all our registrars had passed the membership examination. We therefore considered that the effect of test sophistication on scores would be minimal.

APPENDIX 5

Policy on Guessing, Instructions to Respondents and Scoring Schedule for Questionnaires

Decisions to be taken about the appropriate MCQ scoring schedule and instructions to be given to subjects about responding to options are closely related, each being dependent, to a certain degree, on the test constructor's approach to the question of whether or not candidates should be instructed to guess when they are unsure of the correct response to an option. Although the following discussion deals with these three aspects sequentially, in practice the decisions on each are taken in parallel.

The question of guessing has been widely discussed in the literature (see, for example, the discussions of Cross and Frary, 1977; Harden et al, 1976; Lord, 1975; Palva and Korhonen, 1973; Rowley and Traub, 1977; Wood, 1977). However, only a brief exposition will be given here since it is argued that the question itself is not crucial given the role of our multiple choice questionnaires which is to discriminate between groups (students and registrars) rather than individuals and compare those groups not on total scores but on the relationship between scores on each of the four sections of each questionnaire.

Rowley and Traub (1977) accurately summarise the controversial area:

"... the question of whether to encourage or to discourage guessing has yet to be satisfactorily resolved. The dilemma one faces in connection with this decision may be summarised as follows: if one encourages students to answer all questions, whether they know the answer or not, a source of random variance is introduced which decreases both reliability and validity; on the other hand, if one attempts to discourage students from guessing, it is apparent that some students will comply to a greater extent than others, causing the test results to be contaminated by personality factors which the test was not intended to measure". (pp. 16-17)

A number of workers suggest that candidates should be encouraged to guess (for example, see Crehan et al, 1976; Harden et al, 1976), although suggested instructions vary from answering every item to

answering only those items on which candidates feel themselves to have at least partial knowledge. Arguments in favour of guessing may be substantiated by studies showing that if candidates are forced to guess, where previously they had chosen not to respond, their total score increases (for example, Cross and Frary, 1977). An alternative, or additional, argument in favour of guessing is expressed by Harden et al (1976):

"... since the random error introduced by instructions to guess is less damaging to the meaning of the score than the error introduced by varying tendencies to guess ..., there are very good arguments in favour of discontinuing the 'don't know' option in multiple choice examination papers". (p. 32)

In addition, the 'don't know' option might hide partial knowledge and candidates of equal partial knowledge might not have equal willingness to commit themselves to an answer. Sanderson (1973) also comments on increased variability due to the 'don't know' option. Arguments against encouraging guessing tend to be based on judgments that it is psychologically wrong or 'unprofessional' to encourage guessing (for example, see Clarke et al, 1976; Lennox, 1967; Rowley and Traub, 1977) that, particularly in medicine, it is important for a candidate to recognise when he does not know the answer to a problem (Harden et al, 1976), and that the use of a guessing penalty improves the reliability and validity of the scores obtained on the test (see Rowley and Traub, 1977).

As an area of research, the question of encouragement to guess or not is made difficult to answer by the seeming lack of independence of the three main variables (scoring schedule, instructions to subjects and guessing) and the number of different forms that each of these variables may take (for example, scoring schedules may have different formulae for penalty or counter-marking, instructions to subjects across independent research studies take differing forms, guessing may be advised more or less strongly or on the basis of partial or no knowledge). It is extremely difficult, therefore, to interpret and compare research findings.

The decision about guessing, therefore, was largely made on grounds other than those cited above. Firstly, as already indicated, the

purpose was not to compare one candidate's score with that of another, but was to compare one group of candidates with another group primarily in terms of the relationship found within each group between scores on each of the four sections of each questionnaire. Where the groups are compared in terms of absolute scores, it is reasonable to assume (although the assumption is not tested) a similar distribution of personalities across groups, therefore a similar range and distribution of test taking strategies. Secondly, the decision was taken in terms of the subjects' test taking experience. The 'don't know' option is still available in most MCQ's used by British medical schools, and the Royal College of Physicians. Availability of the 'don't know' option was therefore more likely to be commensurate with the subjects' past experience. Finally, responses to these questionnaires were unspeeded and subjects were not subjected to any of the competitive or examination stresses which may trigger guessing strategies in those with that propensity, although this is not to suggest that guessing behaviour will have been eliminated among the subjects.

The instructions to candidates, then, were as follows:

"In the following questions, put a tick (✓) if the statement is correct; put a cross (X) if the statement is incorrect. Leave the space blank if you do not know the answer. Do not guess".

No further information about, for example, the scoring schedule was considered necessary or desirable, since the questionnaires were to be implemented as teaching tools, not as testing instruments, and maximisation of scores was not as important a criterion as consistent test taking strategy for each individual across all four sections of the questionnaire.

Having decided on encouragement not to guess, the selection of a scoring schedule becomes less problematical, since scoring schedule and instructions to candidates on guessing are complementary.

The basic choice of scoring schedule is between number right scoring and formula scoring systems. Number right scoring gives the candidate credit for each correct response, therefore test instructions

should inform the candidate that to answer every item is the best strategy, since no penalty is given for incorrect responses. Formula scoring, however, penalises wrong answers by giving a counter mark, therefore test instructions should inform the candidate that he should not guess on those items to which he does not know the answer. Lord (1975) suggests that the candidate should also be advised to answer the item whenever he has any valid partial information to guide him. We did not offer specific guidance on this point, considering that each subject's own judgment of whether or not he 'knew the answer' would not be assisted by drawing his attention to the concept of partial information about which he would also have to make a similar judgment. One candidate's idea of partial information may be another's of knowing the answer and yet another's of not knowing the answer.

Given the other characteristics of the questionnaires, then, a formula marking schedule was the only one appropriate. Lord (1975) quotes the usual formula score as

$$R - \frac{(\text{number wrong})}{A - 1}$$

where R is the number of right answers and A is the number of choices per item. The MCQ's may each be taken as a series of separate true/false questions (Harden et al, 1976; Jolly, 1976) and so in this formula (A-1) is zero. The formula used therefore is

$$R - (\text{number wrong})$$

which is equivalent to awarding +1 for a correct response, -1 for an incorrect response and zero for no response. This schedule has been very widely used in Britain (Buckley-Sharp and Harris, 1971; Fleming et al, 1976; Harden et al, 1976) although some modifications have been suggested, such as confidence testing (Palva and Korhonen, 1973; Rothman, 1969) or greater weighting of the counter mark (Lennox, 1967). Whitby (1977) gives a short review of advantages and disadvantages of no penalty and penalty marking systems, attempts at correction for guessing, the use of the 'don't know' option and confidence weighting. Although his review is far from comprehensive, it gives indication of the breadth of thinking about these questions within medical education alone.

APPENDIX 6

The Intraclass Correlation Coefficient: Nature, Purpose and Formula

In essence, the intraclass correlation coefficient (R) is an index of similarity of responses given by respondents in a single class. Haggard (1958) defines the index as follows:

"The coefficient of intraclass correlation is the measure of the relative homogeneity of the scores within the classes in relation to the total variation among all scores in the table. Thus maximal positive correlation exists when all the intraclass scores are identical and the scores only differ from class to class. As the relative heterogeneity of intraclass scores increases, the computed value of R will decrease; maximal negative correlation exists when the heterogeneity of the intraclass scores is maximal and all the class means are the same". (Haggard 1950, p6)

It will be clear, then, that R is closely related to F. Haggard (1958) shows that R and F are functionally related and that the relationship is monotonic. However, there are two important differences between R and F. Firstly, R is a test statistic, having no theoretical upper limit and whose meaning is a function of the number of degrees of freedom, whereas F may be used as a general descriptive statistic whose maximal value is +1 and whose magnitude is independent of the number of associated degrees of freedom. Secondly, F and R enable one to answer different questions: F pertains to questions of difference, whereas R pertains to questions of similarity. However, the coefficient of intraclass correlation is phrased in terms of a ratio of components of variance which can be estimated directly from the usual analysis of variance tables. The formula for estimating R may be written

$$R = \frac{BCMS - WMS}{BCMS + (k-1)WMS}$$

where BCMS is the between-classes mean square

WMS is the within-classes mean square

k is the number of members per class in the sample

The level of significance of R is identical with that of the corresponding F (ie BCMS/WMS), although the F table is generally appropriate only for one-tail tests, that is where $BCMS > WMS$.

Although Lewy (1973) recommends the use of the intraclass correlation coefficient in the selection of items for a test, we are using the statistic as Haggard (1958) demonstrates it, in analysis of the test as a whole. Given that test homogeneity is determined, this is an appropriate analysis, not even requiring the separate calculation of R for each section of the questionnaire.

APPENDIX 7

Calculation of the Intraclass Correlation Coefficient (R) for
Endocrinology and Neurology Questionnaires Separately

(N.B. All figures have been rounded to one decimal place for these tables. The original calculation worked to 3 decimal places).

1. Questionnaire in Endocrinology

<u>Class</u>	<u>Number</u>	<u>Σx</u>	<u>Σx^2</u>
Registrars	35	5149	768615
Students	35	3520	376600
		<u>Σ 8669</u>	<u>Σ 1145215</u>

$$\text{Correction Term (C)} = \frac{8669^2}{70} = 1073593.7$$

$$\text{Total SS (69df)} = 1145215 - 1073593.7 = 71621.3$$

$$\text{Between SS (1df)} = \frac{(5149^2 + 3520^2)}{35} - C = 37909.2$$

$$\text{Within SS (68df)} = 33712.1$$

$$\text{BCMS} = 37909.2$$

$$\text{WMS} = 495.8$$

$$\begin{aligned} \text{R} &= \frac{\text{BCMS} - \text{WMS}}{\text{BCMS} + (k-1)\text{WMS}} \\ &= 0.68 \end{aligned}$$

Statistical Significance of R

$$\begin{aligned} \text{F corresponding to R} &= \frac{\text{BCMS}}{\text{WMS}} = \frac{76.46}{1} \text{ (df. 1,68)} \\ &\quad (p < .001) \end{aligned}$$

2. Questionnaire in Neurology

<u>Class</u>	<u>Number</u>	<u>Σx</u>	<u>Σx^2</u>
Registrars	35	5288	809364
Students	35	4007	475865
		<u>Σ 9295</u>	<u>Σ 1285229</u>

$$\begin{aligned}
 \text{Correction Term (C)} &= \frac{9295^2}{70} = 1234243.2 \\
 \text{Total SS (69df)} &= 1285229 - 1234243.2 = 50985.8 \\
 \text{Between SS (1df)} &= \frac{(5288^2 + 4007^2)}{35} - C = 23442.4 \\
 \text{Within SS (68df)} &= 27543.4 \\
 \text{BCMS} &= 23442.4 \\
 \text{WMS} &= \frac{27543.4}{68} = 405.0 \\
 \text{R} &= \frac{\text{BCMS} - \text{WMS}}{\text{BCMS} + (k-1)\text{WMS}} = \underline{0.62}
 \end{aligned}$$

Statistical Significance of R

$$F \text{ corresponding to R} = \frac{\text{BCMS}}{\text{WMS}} = 57.9 \text{ (df. 1, 68)}$$

$$(p < .001)$$

APPENDIX 8

Calculation of the Test Difficulty Index

Test difficulty is defined by Hubbard and Clemens (1961) as "the ratio of the average number of correct responses, or test mean score, to the maximum number of correct responses". It may be expressed thus:

$$\text{Test difficulty} = \frac{\text{Test Mean}}{\text{Number of items in test}}$$

As with measures of item difficulty, the lower the value yielded, the more difficult the test. We would expect test difficulty values to be relatively high for registrars and relatively low for students.

APPENDIX 9

Calculation of the Phi Coefficient

Values of the phi coefficient were obtained by taking the upper 25 scoring registrars and the lower 25 scoring students and calculating the percent in each group correct for each item of the questionnaire.

The value of ϕ was then read from a chart prepared for estimating the phi coefficient (given in Hubbard and Clemans, 1961) except in borderline cases when ϕ was calculated according to the formula.

$$\phi = \frac{P_u - P_l}{2 \sqrt{pq}}$$

where P_u = per cent passing item in the upper
50 per cent of scorers ie registrars

P_l = per cent passing item in lower
50 per cent of scorers ie students

P = mean of P_u and P_l

$q = 1 - p$

APPENDIX 10

Values of the Phi Coefficient (ϕ) for the Questionnaires in
Endocrinology and Neurology

ENDOCRINOLOGY			NEUROLOGY		
Item	ϕ	Significance level	Item	ϕ	Significance level
Section A			Section A		
1	0.45	$p < .05$	1	0.35	$p < .05$
2	0.15	N.S.	2	0.75	$p < .05$
3	0.20*	N.S.	3	0.75	$p < .05$
4	0.40	$p < .05$	4	0.45	$p < .05$
5	0.30	$p < .05$	5	0.50	$p < .05$
6	0.40	$p < .05$	6	0.60	$p < .05$
7	0.35	$p < .05$	7	0.70	$p < .05$
8	0.50	$p < .05$	8	0.10	N.S.
9	0.30	$p < .05$	9	0.30	$p < .05$
10	0.30	$p < .05$	10	0.50	$p < .05$
11	0.35	$p < .05$	11	0.50	$p < .05$
12	0.70	$p < .05$	12	0.35	$p < .05$
Section B			Section B		
1	0.60	$p < .05$	1	0.29*	$p < .05$
2	0.65	$p < .05$	2	0.65	$p < .05$
3	0.30	$p < .05$	3	0.60	$p < .05$
4	0.10	N.S.	4	0.40	$p < .05$
5	0.20	N.S.	5	0.45	$p < .05$
6	0.30	$p < .05$	6	0.45	$p < .05$
7	0.29*	$p < .05$	7	0.30	$p < .05$
8	0.60	$p < .05$	8	0.60	$p < .05$
9	0.25*	N.S.	9	0.65	$p < .05$
10	0.55	$p < .05$	10	0.35	$p < .05$
11	0.40	$p < .05$	Section C		
12	0.40	$p < .05$	1	0.50	$p < .05$
Section C			2	0.30	$p < .05$
1	0.55	$p < .05$	3	0.30	$p < .05$

2	0.50	$p < .05$	4	0.50	$p < .05$
3	0.40	$p < .05$	5	0.20	N.S.
4	0.31*	$p < .05$	6	0.30	$p < .05$
5	0.50	$p < .05$	7	0.60	$p < .05$
6	0.60	$p < .05$	8	0.75	$p < .05$
7	0.50	$p < .05$	9	0.70	$p < .05$
8	0.60	$p < .05$	10	0.40	$p < .05$
9	0.22*	N.S.	11	0.65	$p < .05$
10	0.50	$p < .05$	12	0.75	$p < .05$
11	0.70	$p < .05$	13	0.65	$p < .05$
12	0.70	$p < .05$	14	0.50	$p < .05$
13	0.60	$p < .05$	15	0.50	$p < .05$
14	0.70	$p < .05$			
15	0.80	$p < .05$			

For each questionnaire:

Number of students: 25

Number of registrars: 25

Value of ϕ at 5 per cent significance level for 50 subjects: 0.28

Values of ϕ marked with an asterisk were calculated according to the formula described in Appendix 9.

APPENDIX 11

The Split Half Test of Reliability: Method and Correction Factor

To apply the split half technique, the test is divided into two halves (for example, by assigning odd numbered items to one half and even numbered items to the other half) and a correlation coefficient calculated between them. A correction factor should be applied for length of test, since this may have some influence on its reliability. The reliability coefficient for a whole test may be estimated using the Spearman-Brown formula (Ferguson, 1966):

$$r_{xx} = \frac{2r_{hh}}{1+r_{hh}}$$

Where r_{xx} is the reliability coefficient for the whole test

r_{hh} is the reliability of the half test

The statistical significance of the resultant value of r may be tested as described by Ferguson (1966), taking the null hypothesis as:

$$H_0 : r = 0$$

with $N - 2$ degrees of freedom

APPENDIX 12

Notes on the Interpretation of Regression Summary Tables

R represents the correlation between the criterion and the weighted sum of the predictors.

R^2 (the variance of the regression function) is obtained by multiplying each correlation of a predictor with the criterion by its corresponding regression coefficient and summing these products. This regression function examines the inter-relationship of the set of variables and describes how well that set of variables predicts the criterion variable. That is to say that the increase in R^2 tells us how much of the variance on the criterion variable is explained by the variables as they are added to the regression equation.

We must be circumspect in interpreting the regression function. We may speculate about relative contributions of the predictor variables but must remember that the obtained prediction results from a system of predictors in which the elements interact in a complex fashion.

A final point to be remembered is that if two variables have a fairly high correlation with the criterion and a low correlation with each other, both measure different aspects of the criterion and both will contribute substantially to prediction. But if two variables have a high correlation with each other, they are measures of much the same thing, and the inclusion of both, instead of either one or the other, will contribute little to the prediction achieved.

APPENDIX 13

Discriminant Analysis as Applied to the Questionnaires in Endocrinology and Neurology

In applying a discriminant analysis the intention is to distinguish statistically, between the two groups of subjects, and to consider the nature of that distinction in order to infer some distinguishing features of the groups. The four sections of each multiple choice questionnaire represent a collection of variables that measure characteristics on which the groups might be expected to differ. Discriminant analysis optimises these variables in terms of their power to discriminate between the groups. Klecka (1975) describes this operation thus:

"The mathematical objective of discriminant analysis is to weigh and linearly combine the discriminant variables in some fashion so that the groups are forced to be as statistically distinct as possible. In other words, we want to "discriminate" between the groups in the sense of being able to tell them apart".

In many cases, no single issue will perfectly differentiate between groups, but by taking several issues and mathematically combining them, a single dimension may be found on which one group is clustered at one end, and the other group at the other end.

Tatsuoka (1971) states that:

"... the problem of studying the direction of group differences is, equivalently, a problem of finding a linear combination of the original predictor variables that shows large differences in group means. Discriminant analysis is a method for determining such linear combinations".

Alternatively, Cooley and Lohnes (1971) describe the derivation of discriminant functions slightly differently:

"In research studies involving several samples from different populations located at different places in a multivariate measurement space, but assumed to be samples from populations having a common dispersion, it can be very interesting to locate the best reduced-rank model for parsimoniously but effectively describing

the measured differences of the groups. In multiple discriminant analysis the samples are projected from their places in the complete measurement space into a suitable sub-space".

Once the discriminant functions have been derived, these may be used for two purposes, either for analysis, or for classification. A full discussion of the latter approach is given in Hope (1968), but, for the present study, the former technique of analysis predominates, while classification is only used as an indication of the adequacy of the discriminant functions derived.

Two alternative methods of discriminant analysis are available, direct analysis and stepwise analysis. Using the direct method, all independent variables are entered into the analysis concurrently. The discriminant functions are created directly from the entire set of variables, regardless of the discriminating power of each. Using the stepwise method, all independent variables are entered into the analysis on the basis of their discriminating power. Often, the full set of variables contains excess information about the group differences, or perhaps some of the variables may not be very useful in discriminating among the groups. By sequentially selecting the 'next best' discriminator at each step, a reduced set of variables will be found which is as good as, or better than, the entire set. Both forms of analysis were undertaken in the present study and were computed using the SPSS programme 'DISCRIMINANT' on the ULCC CDC6600 computer. The stepwise procedure used Rao's V as the selected criterion.

Rao's V, the criterion according to which best discriminating variables are next selected in turn, is a generalised distance measure. The variable selected is the one which contributes the largest increase in V when added to the previous variables. This amounts to the greatest overall separation of groups. A variable is considered for selection only if its partial F ratio is larger than a specified value. This ratio measures the discrimination introduced by the variable after

taking into account the discrimination achieved by the other selected variables. Each subsequent discriminating variable is selected as the variable best able to improve the value of the discrimination criterion in combination with previously selected variables. The points made in the final two paragraphs of Appendix 12, therefore, also apply to the interpretation of the results of stepwise discriminant analysis.

APPENDIX 14

The Scheffé Method as Applied to Data from Questionnaires in
Endocrinology and Neurology, after One-Way Analysis of Variance

The method as applied is described by Ferguson (1966).

1. Calculate F ratios using the formula:

$$F = \frac{(\bar{X}_1 - \bar{X}_2)^2}{s_w^2 (n_1 + n_2) / n_1 n_2}$$

2. From tables, obtain a value of F at the required significance level for $df_1 = K-1$ and $df_2 = N-K$.
3. Calculate F' , which is $K-1$ times the required F.
4. Compare the values of F and F' . For any difference to be significant at the required level, F must be at least equal to F' .

Note:

Ferguson (1966) suggests that concern may attach to the fact that the Scheffé procedure is more rigorous than others, and will lead to fewer significant results. Scheffé's own recommendation to employ a less rigorous significance level, such as .10, is quoted. This level is accordingly adopted here, although inspection of results shows that the same pattern of statistically significant and non-significant results would emerge if a significance level of .01 were adopted.

APPENDIX 15

Procedures for Normalisation of Raw Scores

All scores were normalised as percentages. For each section of each questionnaire the theoretical range of scores may differ. These values are as follows:

<u>Questionnaire</u>	<u>Section</u>	<u>Theoretical Minimum</u>	<u>Theoretical Maximum</u>
Endocrinology	A	-60	+60
	B	-60	+60
	C	-75	+75
	D	0	+6
Neurology	A	-60	+60
	B	-50	+50
	C	-75	+75
	D	0	+15

For each section D, normalisation to percentage values was achieved by dividing the gained score by the theoretical maximum and multiplying by 100, thus:

$$100 \text{ (Gained score/Theoretical maximum score)}$$

For sections A, B and C, the procedure for normalisation of raw scores took into account the negative sign of the theoretical minimum scores. In those cases, then, the theoretical maximum score was added to the gained score, and this sum divided by twice the theoretical maximum score and multiplied by 100 thus:

$$100 \text{ (Gained score + Theoretical maximum score) / 2 (Theoretical maximum score)}$$

This effectively relocates the theoretical minimum score to zero, and expresses the gained score as a percentage of the new theoretical maximum score which is now twice the value of the original theoretical maximum score.

APPENDIX 16

Discriminant Analysis Results and Discussion

Four discriminant analyses were performed, one direct and one stepwise for endocrinology and neurology questionnaires separately. The results of each of these analyses will be reported separately.

A. Direct Analysis - Endocrinology

Table A shows the values of Wilks' lambda (U-statistic), associated F values and levels of significance for each of the four variables (sections A, B, C and D). These represent the results of an

<u>Table A. Direct Discriminant Analysis (Endocrinology). Values of Wilks' Lambda (U-statistic)</u>			
<u>Variable</u>	<u>Wilks' Lambda</u>	<u>F value (d.f.1,68)</u>	<u>Level of significance</u>
Section A	0.7058	28.35	$p < .01$
Section B	0.6834	31.50	$p < .01$
Section C	0.5463	56.47	$p < .01$
Section D	0.8325	13.68	$p < .01$

analysis of dispersion to test the differences between the means of the two groups to determine whether or not we are correct in supposing that they are not drawn from one and the same population. Although we know that the two groups are drawn from different populations of persons, the analysis of dispersion concerns populations of measures or characteristics of persons, not a population of persons qua persons. The characteristics here selected might include no measures on which the groups actually differ. Inspection of Table A shows, however, that our samples are drawn from different populations of characteristics. We may therefore proceed to the results of the discriminant analysis proper.

The maximum number of discriminant functions possible either is one less than the number of groups or is equal to the number of discriminant variables if there are more groups than variables. In this case, therefore, we may expect one discriminant function only. Table B shows the values of the canonical correlation and Wilks' lambda for that one discriminant function.

Table B. Direct Discriminant Analysis (Endocrinology). Values of Canonical Correlation and Wilks' Lambda.

<u>Canonical correlation</u>	<u>Wilks' lambda</u>	<u>Chi-square</u>	<u>df.</u>	<u>Level of significance</u>
0.7354	0.4592	52.14	4	$p < .000$

From the information provided in Table B, we may judge the importance of the derived discriminant function. The square of the canonical correlation indicates the proportion of variance in the discriminant function explained by the two groups. This value is 0.5408, or 54 per cent. This is a relatively low value, indicating that the direct combination of independent variables does not produce a highly efficient discriminant function.

The value of Wilks' lambda (0.4592) reflects this conclusion. Wilks' lambda is an inverse measure of the discriminating power in the independent variables. The smaller the value of lambda, the greater the discriminating power. The present value of lambda does not suggest a very powerful discriminant function, although it does have some degree of utility. The associated value of chi-square with its level of significance at less than 0.1 per cent suggests a low probability of the derived value of lambda occurring due to the chance of sampling alone.

Table C shows the values of the standardised discriminant function coefficients associated with each independent variable. Ignoring the sign,

each coefficient represents the relative contribution of its associated variable to the discriminant function. The sign denotes whether that contribution is positive or negative. We may use these values to 'name' the discriminant function by identifying its dominant characteristic. In this case section C (selecting and testing possible diagnoses) defines the function, with some contribution from section A (mastery of factual knowledge). To some extent these may represent operations of seemingly less cognitive complexity than those tested in section B (interpretation of symptoms and signs) and section D (formulating a diagnosis).

Table C. Direct Discriminant Analysis (Endocrinology). Values of Standardised Discriminant Function Coefficients.

<u>Variable</u>	<u>Value</u>
Section A	-0.5205
Section B	-0.1503
Section C	-1.0184
Section D	-0.0597

A final test of the adequacy of the derived discriminant function may be made with reference to the according classification of cases. By classifying the cases used to derive the function in the initial instance and comparing predicted with actual group membership, we can measure empirically the success in discrimination by observing the proportion of correct classifications. If a large proportion of misclassifications occurs, then we can conclude that the variables selected are poor discriminators.

Table D shows that 80 per cent of the subjects were classified correctly using the derived discriminant function. More registrars were correctly classified than were students (31 as opposed to 25).

Almost one third of the students (10) were misclassified. This finding follows on from and gives further support to the previous indications that the variables selected do not, when combined, produce a highly powerful discriminant function, although that function does have an appreciable degree of discriminatory power.

Table D. Direct Discriminant Analysis (Endocrinology).
Classification of Cases.

Actual Group	Predicted Group Membership	
	Students	Registrars
Students	25 (35.7 per cent)	10 (14.3 per cent)
Registrars	4 (5.7 per cent)	31 (44.3 per cent)

B. Stepwise Analysis - Endocrinology

Although, in essence, this analysis produced the same ultimate results as the direct method, it does furnish us with information about the intermediate stages culminating in those same conclusions and thus enables a deeper and more accurate understanding of the derivation of the function and the relative contribution of each variable to that final value.

Given Table A, we may immediately inspect the summary table of the stepwise discriminant analysis, presented in Table E.

Table E. Stepwise Discriminant Analysis (Endocrinology).
Summary Table.

Variable entered	Wilks' lambda	Level of significance	Rao's V	Change in Rao's V	Level of significance
Section C	0.5463	$p < .000$	56.47	56.47	$p < .000$
Section A	0.4630	$p < .000$	78.87	22.40	$p < .000$
Section B	0.4598	$p < .000$	79.89	1.02	$p < .313$
Section D	0.4592	$p < .000$	80.08	0.19	$p < .668$

The results of the stepwise analysis reflect exactly the findings of the direct analysis, and may be so interpreted in conjunction with Tables B and C. We have further evidence that sections C and A are instrumental in the discriminant function, while sections B and D add very little to the value of Rao's V, and subtract little from the value of Wilks' lambda. We must bear in mind that each successive variable is selected given the variables already contributing to the function; we may therefore conclude only that sections C and A form a reduced set of variables which is almost as good as the entire set in discriminating between the two groups.

C. Direct Analysis - Neurology

Table F shows the values of Wilks' lambda (U-statistic), associated F values and levels of significance for each of the four variables. It can be seen that our samples are drawn from different populations of characteristics. We may therefore proceed to the results of the discriminant analysis proper.

<u>Table F. Direct Discriminant Analysis (Neurology). Values of Wilks' Lambda (U-statistic)</u>			
<u>Variable</u>	<u>Wilks' lambda</u>	<u>F value (df 1,68)</u>	<u>Level of significance</u>
Section A	0.5206	62.63	$p < .01$
Section B	0.7248	25.81	$p < .01$
Section C	0.6791	32.13	$p < .01$
Section D	0.6621	34.70	$p < .01$

Table G shows the values of the canonical correlation and Wilks' lambda for the discriminant function. The value of the squared canonical correlation is 0.5726, which indicates that 57 per cent

Table G. Direct Discriminant Analysis (Neurology). Values of Canonical Correlation and Wilks' Lambda.

<u>Canonical correlation</u>	<u>Wilks' lambda</u>	<u>chi-square</u>	<u>df</u>	<u>Level of significance</u>
0.7567	0.4275	56.94	4	$p < .000$

of the variance in the discriminant function is accounted for by the two groups. This is a relatively low value, although slightly higher than that associated with the endocrinology study, and indicates that the direct combination of the independent variables does not yield a highly efficient discriminant function. The statistically significant value of Wilks' lambda (0.4275) reflects this conclusion although, as with the discriminant function associated with the endocrinology data, the function does have some degree of power.

Table H shows the values of the standardised discriminant function coefficients associated with each variable. The dominant variable associated with the function is section A, followed by section B and almost equivalent contributions from sections C and D. We see here then, a different picture from that drawn by the same coefficients on the endocrinology data, where section C is predominant. However, again it is a variable of seemingly less cognitive complexity which contributes most to the discriminant function.

Table H. Direct Discriminant Analysis (Neurology). Values of Standardised Discriminant Function Coefficients.

<u>Variable</u>	<u>Value</u>
Section A	-1.2512
Section B	0.6170
Section C	-0.5290
Section D	-0.5273

Considering the classification of cases on the basis of the derived discriminant function, Table J shows that 85.7 per cent of the subjects were classified correctly. As with the endocrinology study, more registrars are correctly classified than are students (32 as opposed to 28). We may conclude that, although the derived function is not highly powerful, it does have an appreciable degree of discriminatory power in terms of the classification of cases demonstrated.

Table J. Direct Discriminant Analysis (Neurology).

Classification of Cases.

Actual Group	Predicted Group Membership	
	Students	Registrars
Students	28 (40 per cent)	7 (10 per cent)
Registrars	3 (4.3 per cent)	32 (45.7 per cent)

D. Stepwise Analysis - Neurology

Again, this analysis produces the same ultimate result as the direct method, but furnishes us with further information about the derivation of the function. Given Table F, we may immediately inspect the stepwise analysis summary table, given in Table K.

Table K. Stepwise Discriminant Analysis (Neurology).

Summary Table.

Variable entered	Wilks' lambda	Level of significance	Rao's V	Change in Rao's V	Level of significance
Section A	0.5206	$p < .000$	62.63	62.63	$p < .000$
Section C	0.4800	$p < .000$	73.68	11.05	$p < .001$
Section D	0.4558	$p < .000$	81.18	7.50	$p < .006$
Section B	0.4275	$p < .000$	91.08	9.90	$p < .002$

The results of the stepwise analysis do not exactly reflect the findings of the direct analysis, since section C rather than section B is selected after section A. This must reflect a high correlation between sections A and B, presuming section A to account for those parts of section B with which it is correlated. A high correlation between sections A and B would also imply a low partial correlation which relegates section B to being selected as the fourth rather than second variable in the discriminant function. Tables 7.3 and 7.4 give support to this conclusion, showing statistically significant correlations between sections A and B on the neurology questionnaire for both students and registrars.

Reflecting the results of the stepwise analysis for endocrinology, we may conclude that sections A and C form a reduced set of variables which is almost as good as the entire set in discriminating between the two groups. However, the pattern of reduction in the value of Wilks' lambda and parallel increase in the value of Rao's V is rather unlike that shown for endocrinology in Table E. For neurology, section A is the most instrumental variable, whereas for endocrinology section C is most important. The increases in Rao's V upon selection of the second variable are also markedly different. Proportionately, the second variable entered in neurology accounts for a much smaller increase in Rao's V than does the second variable entered in the endocrinology study. In addition, the third and fourth variables selected for neurology play a greater part in the discriminant function than do the parallel variables for endocrinology. However, these differences are predominantly a matter of degree. It must be concluded that in both endocrinology and neurology sections A and C form a reduced set of variables which is almost as good as the entire set in discriminating between the two groups. However, in endocrinology, section C is predominant, whereas in neurology section A is predominant.

E. Endocrinology

Both the direct and discriminant analyses show the same results, we shall therefore consider them jointly. The four sections of the

questionnaire produce a discriminant function of value, which correctly classifies 80 per cent of our subjects. The function, however, is less successful in classifying students than registrars. In addition the function derived has relatively poor values on other indices of power (square of the canonical correlation = 0.5408; Wilks' lambda = 0.4592).

From these results we may conclude that the variables measured in the questionnaire do not constitute a highly appropriate or complete set if these are to be used as the basis of discriminating between students and registrars. If this conclusion is unacceptable, the only alternative interpretation is that the groups are, in reality, not discriminable beyond the attained level of efficiency. At both empirical and commonsense levels, this latter explanation of the results would seem untenable. We must therefore support the conclusion that our present set of variables is incomplete in terms of discriminating between the two groups of subjects.

The results of the discriminant analysis of the endocrinology data best suggest that processes and skills other than those here tested may add to the efficiency of the function obtained.

F. Neurology

Since the results of the direct and stepwise analyses are complementary, they will not be treated separately. The derived discriminant function correctly classifies 85.7 per cent of the subjects, which is slightly better than the similar index for endocrinology, but the values of Wilks' lambda (0.4275) and the square of the canonical correlation coefficient (0.5726) are relatively poor.

From the above indices, we may conclude that the variables which make up the four sections of the neurology questionnaire do not constitute a highly complete or appropriate set from which to derive an efficient discriminant function. As with endocrinology, the alternative interpretation of non-discriminability must be rejected.

We conclude that the results of the discriminant analysis of the neurology data suggest that processes and skills other than those here tested may add to the efficiency of the function obtained.

G. Comparison of Endocrinology and Neurology

Two points of comparison are immediately obvious. Firstly, in both endocrinology and neurology there are variables other than those tested which might yield a more efficient discriminant function. This finding was expected. In both cases, the untested variable might be identified as the dynamic, active (and interactive), clinician-generated aspect of the diagnostic problem solving process. Secondly, on both cases sections A and C form a reduced set of variables which is almost as good as the entire set in discriminating between students and registrars.

In the same reduced sets of variables, we also find a point of considerable contrast between endocrinology and neurology. For endocrinology, the discriminant function is defined by section C (selecting and testing diagnostic possibilities), whereas, for neurology, the function is defined by section A (mastery of factual knowledge). As emphasised above, however, these variables are weighted and do not necessarily pinpoint the areas of greatest difference between the groups when unweighted.

In conclusion, the discriminant analysis has confirmed a greater breadth of the diagnostic process and of skills inherent in the study and practice of endocrinology and neurology than is measured by the multiple choice questionnaires. This is as expected.

APPENDIX 17

Example of Experimenter's Notes Taken During Simultaneous Viewing
of a Patient - House Officer Encounter

<u>Houseman</u>	<u>Patient</u>
Name ?	-
Age ?	59.
Married ?	Yes.
Job ?	Motor Mechanic.
What brought you in ?	Cough, SOB.
How long ?	6/12.
Getting worse ?	Both cough and SOB.
Phlegm ?	In a.m., excess.
Colour ?	White.
Smoker ?	2 cigars daily.
How far could you walk ?	No problem walking.
SOB got worse ?	Yes.
Cough worse ?	Sometimes in a.m.
Not bothered before ?	No.
Coughed blood ?	No.
Ill before ?	Heart strain, year before.
Ops ?	Slipped disc, hernia, thyroid.
Why thyroid ?	Lost weight.
Rh. fever, T.B. ?	No.
How's heart ?	Normal.
Chest pain on exercise ?	No. Tightness since cough.
Wheezing ?	Just tightness.
How many pillows ?	2.
Wake up breathless ?	No.
Ankles swell ?	No.
Leg pain on walking ?	No.
Appetite ?	Good.
Weight ?	Steady.
Heartburn, indigestion ?	Indigestion.
After meals ?	Before.
Bowels ?	No change.

<u>Houseman</u>	<u>Patient</u>
Blood ?	No.
WW ?	OK.
Good stream, etc. ?	Yes.
Headaches, fits, faints ?	Headaches in a.m.
Pins and needles ?	Left arm, down to fingers.
Numbness ?	From left elbow down, tingling in left hand.
How long for ?	6 weeks.
Same period as cough and SOB ?	Coughing when shoulder pain came on.
Weakness ?	Only hand.
Wife ?	OK.
Children ?	One, married.
Parents ?	Mother, 84, heart attack. Father, died World War II.
Sibs ?	7, OK.
T.B. ?	No.
Heart trouble ?	No.
Diabetes ?	No.
Epilepsy ?	No.
Main problem is cough and SOB ?	Yes.
And numbness and tingling ?	Yes.

APPENDIX 18

Transcript of an Account Gathering Session with a Student, with
Experimenter's Notes

Notes

<u>Student</u>	<u>Patient</u>
Name ?	-
Age ?	59
Work ?	Retired.
Problem ?	M.I. July 1975, repeated 1977.

Account

- E. You ask about his name, his age, his work and his present problem, is that your normal way of starting an interview ?
- S. It depends on the patient and on how they appear, if it was somebody who was very ill, no, I would ask them about their present problem and you can usually find out name and age from the end of the bed anyway. But it is important to find out about somebody's work because it may often have a bearing on the situation and also if he's going back to work afterwards; and knowing if he's retired may be important as well.

Notes

-	1/52 ago, 'small infarct'.
What happened ?	Came to coronary unit.
What did you notice wrong ?	Acute discomfort.
Where ?	Pressure, pushing side of heart.

Account

- E. So that was the events of a week ago and him coming into the coronary care unit and you asking for an elaboration of his symptoms and what was wrong, what were you thinking about his problems at the time ?

S. Has he had a coronary or not. What else could he have had, because he went up on to coronary care. It was very likely that he did or, at least, that there was something wrong with his heart-angina or something of that sort. I wanted to find out whether the pain was typical heart pain or if it wasn't.

E. So you're elaborating on his symptoms.

S. Yes.

Notes

Did pain move ?	No.
Arms ?	No.
Faint, sweaty, sick ?	Not in 1975, last was middle of night. Felt sick.

Account

E. You're asking about particular things, fainting, feeling sweaty, feeling sick, pain in his arms.

S. Again, trying to localise whether or not it was typical cardiac pain.

E. And what did it appear to be at that time ?

S. Atypical.

E. So what did you think about it appearing atypical ?

S. Well, there is a variety in the type of heart pain people can have. Old people often don't have any pain at all, and, as he said, they don't know whether or not he'd had a heart attack, and from that history I don't either.

Notes

General health ?	Good.
Pain on exertion ?	No, normal tiredness.
Drugs ?	No, just anticoagulants.
Warfarin ?	-

Account

- E. So that was his general health, pain on exertion, and drugs.
What did you get out of that ?
- S. I wanted to know whether or not he had any angina, which he didn't, because that and a heart attack are part of the same disease. I wanted to know whether he's had any treatment. The sort of thing I might have expected would have been something for hypertension, warfarin or some other anticoagulant, or something for arrhythmia or congestive heart failure. Something like that, I'm just trying to find out.
- E. So you're looking for clues as to what his problem might have been, it being atypical in the way that you say.
- S. I was looking for clues to see whether he has any other heart frailty really; to find out whether he is hypertensive, which would make ischemic heart disease more likely, and that sort of thing. Whether he has any residual damage, whether he has other lesions or heart failure, or arrhythmias, and it seems that he doesn't have any really.

Notes

- | | |
|------------------|-----------------------------|
| General health ? | Good. |
| Why retired ? | Nothing to do with health. |
| Stress ? | No. |
| Chest trouble ? | Cough, 20 cigarettes a day. |

Account

- E. You talked about why he retired and then asked about chest trouble.
- S. His retirement might have been because of heart trouble; even though he talked about pain in 1975 it might have been because it was a particularly stressful situation. It is said by some people that stress causes ischemic heart disease.
- E. You went on to ask about chest trouble and cough.

S. This is part of the systemic enquiry that I'm making at the moment, and which I'll continue to find out if there's anything else wrong with him. Fairly routine sort of screening pattern which I do; but again chest is more relevant to him if he had heart failure or something. He could have shortness of breath which I did go on to ask.

E. So at the moment you're thinking about ischemic heart disease, are you ?

S. Well, now I'm going on to find out whether there's anything else, any other clues which need following up. It's still basically accepting the experts' opinion that it's probably heart and that he probably had an infarct, because they've done the investigations, so there's not much else you can do really.

Notes

Haemoptysis ?	No.
Wake up with SOB ?	No.
Appetite ?	Good.
Weight ?	No change.
Bowels ?	No trouble.
Constipation, blood ?	No.
WW ?	OK.
Everything else OK ?	Yes.

Account

E. What are you thinking there ?

S. Anything else that could be relevant to him. Is there anything else that seems relevant. It's not typical heart pain, but I haven't picked anything up on that which could be relevant. We could go through the differential diagnosis of things that might be relevant to epigastric pain, but there wasn't anything there.

E. At this stage, you're thinking that there's something cardiac but you're not precisely sure what.

S. Throughout I've thought that he had atypical heart pain which was probably caused by an infarct, my reasons for that being basically the pain and, more than that, other peoples' opinions who've had longer with him and such like; and on the past history as well, not just the recent episode. In the past he had an infarct, which is making it much more likely that he would have another one this time. But I am looking for other possible causes for some sort of epigastric pain. I'm just trying to exclude other things really to make sure that there isn't any other pathology or any thing else that has come on which has co-existed with it or which may be causing it.

Notes

Better since in ?	Yes.
Drugs ?	Anticoagulant and others.
Easy to rest, restful life ?	Yes.
Past illnesses ?	No.
Diabetes ?	No. Thrombosis, 1963.

Account

E. Anything there ?

S. Nothing new, no.

E. Could you give a brief summary of this patient ?

S. A generally well, 59 year old man, who's had 2 attacks of probable heart pain, in July 1975 and about a week ago, which were probably caused by an infarct.

E. And what things in particular make you say that, make you conclude that ?

S. In someone of his age it's very common, especially being a smoker, and somebody with a probably stressful job. He appears to have no other cause, no other symptoms at all, and, more than these, it's what other people think. If they've been able to do ECG's and relevant blood tests.

- E. Is there anything in his history that might make you doubt your conclusions ?
- S. The atypical pain. The fact that he didn't feel faint, sweaty or nauseated or vomit, although he did a bit. People normally faint or black out, and feel cold and sweaty. The treatment of warfarin or another anticoagulant is quite likely for the first infarct but it wouldn't have suprised me if he'd been on something else as well, something for hypertension, something for arrhythmias, something for heart failure, but he wasn't on any of those.
- E. If you could have examined him, what would you have done ?
- S. I'd have done the routine type of examination, respiratory, cardiovascular, seeing if there was anything wrong with his lungs, to see if there was any pulmonary oedema, and such like signs of heart failure. Otherwise it would have been a very routine examination, expecting to find little murmurs or ninyasastatmus but it would seem unlikely. He's very well and making such a good recovery. Investigations, especially if I'd seen him immediately, obviously one would have a chest x-ray, ECG, CPK, haemoglobin, these would be the immediate things, and depending upon the results of those, especially if the symptoms didn't get better, which they do seem to be. It might seem a good idea to do a barium meal, to see if there was some lesion of the stomach or duodenum; but if he got better and there was no recurrence of pain I wouldn't do that.
- E. Finally, how typical do you think that history was of you ? Was it your normal way of going about things, your normal style ?
- S. It was very similar. I think everything was there in the end, but it might have come in a slightly different order.

APPENDIX 19

Transcript of an Account Gathering Session with a House Officer,
with Experimenter's Notes

Notes

House Officer

Patient

Age ?

41.

When came in ?

2 weeks ago.

Main complaint ?

Blood in urine.

Account

- S. I usually begin by asking them how old they are, whether they are married or not, and then how many times they've been into hospital, whether it's the first time or not, and then I get straight down to it.
- E. This man's got blood in his urine. What did you think ?
- S. Haematuria, and he's also having a problem with his renal tract, somewhere from his kidneys down to the ureters, his bladder or urethra, and I thought of the causes of haematuria.

Notes

How long for ?

3 years ago, for couple of days.
Stopped after 6-7 months, then
came for couple of days.
Stopped. This time, for last
2 months, more regular.

Account

- S. You want to know whether he's passing blood all the time, or whether it's intermittent. You're thinking about the causes of haematuria and whether he's got a stone, let's say, which could be passing down and so giving him blood at some stage,

and nothing at another stage; or if he's got a history of trauma which means he'll bleed for a while and then stop bleeding. So I'm finding out a bit more about his bleeding. Very non-specific/specific, so that particular part of his history narrowed it down a little if anything. It wasn't too helpful, but it's useful to know.

Notes

Blood, bright red or dark ?	Sometimes bright red; in morning dark, passing all the way through.
Mixed with urine?	Yes; sometimes small drop at end.
Clots ?	Yes, sometimes.
Pain in side ?	No.
No kicks ?	No. 6 years ago, bad fall on side. X-ray, said badly bruised ribs.

Account

- E. You started off with what the blood was like; sometimes it was bright red, in the mornings, sometimes dark.
- S. He could have an inflammation or a tear, say, right low down the end of his penis, and if it bled it would look just like fresh blood, just as though you'd cut yourself. So it would be something fairly low down. If it's higher up it's likely to be darker.
- E. He said sometimes it was one and sometimes the other. What about that ?
- S. That didn't help.
- E. You wanted to know if it was mixed with the urine and how it came out, whether it was coming out all the time or just sometimes.

- S. Well, that points to some sort of tear, really. If he's got a urethral tear or something like that he might bleed after he's passed urine; he might find there's blood coming out after he's stopped peeing.
- E. Again, his answer was sometimes it's one thing and sometimes it's the other.
- S. Signs of prostatism. That localises it below the bladder. I didn't think he had prostatism, although he said he had a poor stream, but then, he might have to pass some sort of obstruction. I asked him how many times he went during the day. If he went lots and lots of times it would point to a urinary tract infection of some sort; but he wasn't passing a lot more water or going to the toilet more frequently.
- E. So what are you thinking about him by this stage ?
- S. I thought the problems were in his bladder specifically. I couldn't say, a bladder tumour or a papilloma, I don't know. Sometimes when you see a patient you have fixed ideas and you're out to prove it. I'm like that. If I get some idea I'll stick to it and I'll pursue that. I know I do it, but I feel so confident about it. Quite a lot of the time you're wrong.
- E. But you didn't have any ideas like that about this man.
- S. At that stage I was beginning to feel strongly about his bladder.
- E. You asked him next about pain in his side, whether he'd ever been kicked.
- S. This all still refers to the renal tract. Any history of trauma, or kicking or accidents to the loin will cause haematuria and I only asked him that after I'd found out he had clots. Sometimes you can get them in the right order.

E. He told you about a fall he had six years ago ?

S. Yes, his problems began after that.

E. So here you begin to feel you can locate it a bit more.

S. I think so, yes.

Notes

Burning, stinging ?	No.
Wind in water ?	No. Everything normal, except for blood.
Weak, lethargic ?	No.
Weight loss ?	No, gained if anything.
Other illnesses ?	No.
Operations ?	No.
T.B., Diabetes, Rheumatic	
Fever, Epilepsy ?	No.
Perfectly healthy ?	Yes.

Account

E. That section began by you asking about burning, stinging, wind in his water.

S. The burning and stinging is urinary tract infection which I thought he didn't have because he didn't have any frequency. The pneumaturia, the wind in his water, is a question that I don't usually ask. If they're passing wind in their urine it means they've got a communication, usually with the bowel. It could be due to cancer or something, or diverticular disease, or there are other causes I can't think of at the moment. It's a specific sign, if they say 'yes', it specifically points to the fact that there's something wrong with the bowel. They might have a cancer there, and they often complain of passing bubbles or gas.

E. Were you then going onto something more routine, when you asked if he was weak and lethargic ?

- S. It's more general questions tied up with this. If he's losing blood I want to know how much he's losing, and when you're anaemic you feel breathless and generally very tired, just generally run down. And I didn't know whether he'd been losing enough, I didn't really quantitate it. I didn't ask him how many buckets of blood he was passing, so I specifically asked him whether he was having any effects from blood loss, I don't think he's lost a lot.
- E. And then you asked him about other illnesses and operations.
- S. That's routine. I specifically asked him about TB because TB causes blood in the urine anyhow, although I had no indication that he had TB.
- E. Why did you go onto routine questions at that point ?
- S. I was stuck. I often do that when I run out of questions. I go onto routine ones. It gives me time to think. I was thinking what else it might be then, and I didn't come up with anything, I wasn't really sure.
- E. So is it still just something in the bladder ?
- S. Yes. I don't think I'm going to be able to narrow it down. I'd like to have a look at it. I didn't really pin him down to anything. I could have probably if I'd really laboured the point, got him to commit himself in one way or the other. Patients' interpretations of symptoms are not always what you'd like them to be. If you're passing clots, it means that blood is being allowed to hang around and when he passes urine it comes out there. So that points to some sort of bladder lesion.
- E. And again he said, yes, sometimes ?
- S. Yes, I'm wondering if he has got something in the bladder. He's going to be cystoscoped later on today, so they're obviously thinking something like that.

E. Is that what you're thinking about at the time you were trying to locate it ?

S. Yes, I was trying to locate at which level in the renal tract

E. Were you feeling here that you were beginning to know ?

S. Yes, it was probably his bladder.

Notes

Tablets ?	No.
Anticoagulants ?	No.
Thirsty ?	No.
Difficulty passing water ?	Sometimes, if I've drunk a lot.
Poor stream ?	. If I've drunk a lot.
Dribbling ?	No.
How many times a day ?	Normally 5-6 times.
Middle of night ?	Only at first, not now.

Account

E. You started off that section with tablets, whether he'd been on anticoagulants.

S. If he's on anticoagulants and he's over anticoagulated he'll bleed, he'll bleed from other places as well. Haematuria is a good sign. I just wanted to make sure he wasn't on anticoagulants.

E. Then you asked whether he was thirsty ?

S. That just flipped through my mind. I was thinking of kidney stones again and hypercalcaemia. Hypercalcaemia makes you thirsty and a high calcium level is a good indication of kidney stones.

E. He wasn't thirsty ?

S. No, I didn't expect him to be.

E. And then questions about passing water.

Notes

Mother and father ?	Mother died, stroke, 10 years ago. Father died, Ca lung.
Sibs ?	One brother, alive, well.
Wife ?	Epileptic.
Children ?	3.
Well ?	Yes.
No-one in family with the problem?	No.
No-one with bleeding disorders or kidney trouble, stones ?	No, mother had gall stones.

Account

E. What about that family history ?

S. I was trying to elicit any history of cancer. It's fairly routine but it is relevant in this case, because if you're thinking of a tumour in the bladder or something like that, then there may be a strong family history of cancer. So specifically I always ask whether there's anything relevant. Sometimes it gives you a clue and sometimes it doesn't. His father had cancer of the lung but no one else had cancer. Cancer doesn't run high in the family. His mother had gall stones, but that was specific. There was nothing specific there. But no one in the family had had bleeding disorders

which could cause haematuria. No-one with kidney trouble, so this was a new thing.

Notes

Job ?	Stores manager.
How long ?	1 year. Before, transport Manager.
Appetite ?	O.K.
Weight ?	O.K.
Bowels ?	O.K.
Passing blood ?	No.
Tarry stools ?	No.
Smoke ?	20 a day.
Drink ?	6 pints daily, 5 days a week.
Allergy ?	No.
Drugs ?	No.

Account

E. His job, his appetite, weight, bowels.

S. Those are routine questions. His job can be specific to his condition. People who work in the aniline dye industry get cancer of the bladder. Sometimes I don't bother with the social history, it all depends on how much time I've got. And the other questions are just review of the systems, just generally to see if there's anything else involved apart from this.

E. And was there anything interesting there ?

S. Apart from the fact that he smoked, smokers have a higher risk of cancer than anyone else and he's a fairly heavy drinker I thought. But I don't think that was really the cause of it. I was thinking more about his bladder, the fact that smokers have a higher likelihood of getting cancer.

E. Are you thinking of cancer in particular ?

S. I'm thinking more of it now.

E. Were you thinking of it at the time ?

S. I suppose I was really.

Notes

Chest pain ?	No.
Palpitations ?	No.
Ankle swelling ?	No.
Chest pain on walking ?	No.
Pain in calves ?	No.
Wake up at night gasping ?	No.
Cough ?	No.
Phlegm ?	No. Occasionally white.
Breathless on back ?	No.
Breathless walking uphill ?	No.
Coughed up blood ?	No.
Fits, faints ?	No.
Skin rashes ?	Boils.
Diabetic ?	No.
Since in, what tests ?	None yet. Cystoscope to come.

Account

S. This is all part of my routine. The first set were just cardiovascular symptoms. He's quite fit really. No signs of any heart failure, or angina, or anything. The coughing and breathlessness again, were just his respiratory system. I'm just seeing what his condition is like.

E. How would you summarise this patient?

S. A fit young man who presents with haematuria, query cause, probably in his bladder.

E. If you could have examined him, what would you have done ?

S. I would have examined him fully. I wouldn't expect to find anything on examination, although I would with tests.

E. What tests would you have had done ?

S. He needs to have an IVP. That's the primary procedure. But before you do that, you do a chest X-ray, abdominal X-ray, take blood, full blood count, ESR, urine electrolytes and I'd take it from there.

E. How typical was that of you ?

S. Eighty per cent. Usually I'm in a hurry so I pressurise them a lot more, and that's where you fix in your mind that they've got something wrong, that they've got what you think is wrong. You're in a hurry and you miss out a lot of questions. A lot of the time you get a feel about the patient. If they're good historians then you believe them, but sometimes they don't answer your questions or they don't understand your questions, so you simplify them, and then you only believe half of what they say. When I'm on the ward clerking patients routinely I take a lot longer than that.

APPENDIX 20

Transcript of an Account Gathering Session with a Registrar,
with Experimenter's Notes

Notes

Registrar

Patient

Name ?

-

Age ?

75

What brought you in ?

Stomach pains. Had small drink.
Dr. said indigestion. Tummy
pains. Finished bottle of stuff.
Pain got worse. Dr. said ulcer.
Tablets.

Account

- E. You ask him his name, his age, and what brought him in. Is that your normal way of beginning an interview ?
- S. It's rather artificial, because he's been in hospital for sometime, whereas one would normally see him in casualty or outpatients first.
- E. He tells you about the pain in his stomach and an ulcer. What did you think about it then ?
- S. One does try to discourage them going on and on about what the doctors had told them. But initially he said that he had an epigastric pain, which may well have been indigestion. I didn't think it was necessarily that. It may have been lots of other things in the duodenum.
- E. What types of other things did you think of or didn't you think of anything in particular ?
- S. It could have been something very non-specific. It could have been a peptic ulcer. It could have been gall stones. It could have been due to angina or a heart cause.

Notes

Tell me more about pain, precipitants ?	2-3 hours after food or drink.
Wake up at night ?	Yes.
Anything make it better ?	Tablets.
Milk helped ?	Yes, or biscuit.
Always in same place ?	Yes.
Vomiting ?	Yes.
With pain ?	No. 2-3 times in last month.
Vomiting related to pain ?	Eased it.
What brought up ?	Brown liquid soup.

Account

- E. You're asking him about the pain in particular, what sort of pitch and what's associated with it. What sort of picture are you getting here ?
- S. I'm seeing if there's a peptic ulcer which causes it or something else wrong with the stomach.
- E. You ask him specifically about its precipitants, and he said that it came on 2-3 hours after food or drink. Did that tell you something specific ?
- S. Well, with duodenal ulcer you get pain when you're hungry, and with a gastric ulcer you get pain when you've just had something to eat. But I don't think that's any longer a valid distinction. A gastric ulcer might wake you up in the middle of the night and be relieved by having something to eat or drink.
- E. In fact that was the case, that it woke him up at night and having something to eat eased it.
- S. It was eased by milk, so I thought it might be a duodenal ulcer.
- E. You asked him whether the pain was always in the same place.

- S. If the pain had been due to heart disease it would have been up in the chest. It might have gone up to the jaw, down in the arm. It could have been due to pancreatitis. He might have had a history of pancreatitis, in which case it would radiate to the back. If it had been gall bladder, it might have radiated right across the epigastrium.
- E. So you're asking him where it's located to rule out things.
- S. I didn't rule them out, I left them in abeyance.
- E. You asked about vomiting and its association with the pain.
- S. The fact that vomiting relieved the pain, means that it could be either a gastric or a duodenal ulcer. The other thing one is thinking about, is that he might have a cancer in his stomach or something like that. And with this vomiting one wants to know what he's vomiting up. It might tell you whether he has some form of pyloric stenosis secondary to a cancer, but he's just bringing up a brown fluid that tells me that he's unlikely to have a pyloric stenosis but he may well have a duodenal ulcer or a cancer of the stomach.

Notes

Eating O.K. ?	Yes.
Weight O.K. ?	Losing. Stone in last 6 weeks.
Only vomitted 3 times ?	Yes.
Bowels ?	Constipated.
How many times ?	Waited 3-4 days.
When came on ?	Over 3-4 months ago. Beer didn't cure it.
Motions normal. No blood ?	No, but black. A lot of wind and belching.

Account

- E. Are you getting anything out of this ?
- S. The fact that he's eating well is really neither here nor there, but the fact that he's losing weight could well be significant.

People with an ordinary peptic ulcer don't normally lose weight unless they've got some complication like a pyloric stenosis, which means they can't keep food down. People with stomach ulcers tend to lose weight. He says he's lost a stone in 6 weeks. The fact that he's constipated is probably only relevant in an indirect way. If you bleed into your gut, you probably get a lot of diarrhoea, and you get a lot of blood in the motions. And he said that his motions were black sometimes. He may have been losing blood, but he didn't have diarrhoea, so that suggests that he hasn't lost a lot of blood.

E. Does it suggest anything else ?

S. That his motions were black only suggests that he's been on iron tablets.

E. What's your overall picture of him at the moment? Are you thinking in terms of malignancy ?

S. I'm keeping that in mind. I'm thinking also of duodenal or gastric ulcer.

Notes

Tablets ?	For gout, water.
Aspirin ?	No.
Alcohol ?	2 pints a week.
How much was most ?	3-4 pints daily.
Smoker ?	No.
Serious illnesses ?	Broken leg, bad ankles, 2 hernias, haemorrhoids.

Account

E. Were the tablets of interest ?

S. Only that in some tablets cause ulceration of the stomach. I asked about aspirins specifically. And the other thing is, I'm just looking for other causes of peptic ulceration.

- E. So all those questions mean that you're looking specifically at that possibility.
- S. Yes, that he drinks 2 pints a week and doesn't smoke doesn't tell me anything.
- E. The section ends with you asking about serious illnesses. Does that mean you're beginning a more routine enquiry ?
- S. Yes.
- E. Why did you go onto a routine enquiry then ?
- S. I thought that I couldn't learn any more from the enquiry.
- E. So were you feeling reasonably happy about what was wrong with him ?
- S. I haven't proved beyond all doubt. I knew what investigations I would order, I would do a barium meal, blood count.

Notes

- | | |
|-------------------------------------|-----------------------------------------------------------|
| No past ulcers ? | No. |
| Job ? | Retired 16 years. Chemical engineer. |
| No pain before this year ? | No. |
| Married ? | Not now. |
| Quite well apart from pain ? | Yes. |
| Chest pain ? | MI 3 years ago. Angina 9 years ago.
Bronchial trouble. |
| Pain in chest when walking uphill ? | Can't walk far, because of pain in throat. |
| Walk upstairs ? | Dr. said not to. |
| Ankle swelling ? | Not now. |
| Wake up with SOB ? | No. Just pain. |

Account

- S. This is mostly routine, but if he had ulcers in the past it would be more likely that he had then now and not a cancer or something. If he's been losing blood, he might be anaemic which could cause chest pain and shortness of breath. Or he could have had heart disease or lung disease. It's always worth asking.

- E. He told you that he'd had angina, he'd had a heart attack and couldn't walk uphill.
- S. It's good to have an idea about him and about his general fitness what he can and can't do. It would be important if you were to come to operate on him.
- E. He said he couldn't walk up a hill because he had a pain in his throat.
- S. Yes, angina can give you a pain in the throat.

Notes

- | | |
|------------------|----------------------------------|
| Cough ? | No. |
| WW ? | O.K., get up 2-3 times at night. |
| Blood in water ? | No. |
| Headaches ? | No. |
| Fainting ? | No. |

Account

- S. These are just routine questions that you ask as they come to mind. You would just ask to find out whether he had any malignancy of any kind, or if he coughed up any blood.
- E. If you could have examined this patient, what would you have done and what would you have expected to have found ?
- S. I would see if he was jaundiced, because he might have had biliary colic. Pulse, and blood pressure, and see if he was tender in the abdomen, whether there was any distention. I would probably have done a rectal examination, if his motions were very black, full blood count etc.
- E. What are your conclusions about this patient at the moment ?
- S. I think I'd concluded that he'd had a duodenal ulcer, but I'm a bit worried about his weight loss.
- E. How typical of you was that interview ?
- S. Fairly typical.

APPENDIX 21

Diagnosis Given in Each Patient's Medical Record - Cases Seen by Students

- Patient 1 (i) Thyrotoxicosis (ii) Allergy to ATD
- Patient 2 (i) Subarachnoid haemorrhage (ii) to (iv) Post-gastrectomy/vagotomy diarrhoea, hypoglycaemia, dumping
- Patient 3 (i) Rheumatoid arthritis (ii) Arteritic ulcer
- Patient 4 (i) Asthma (ii) Hypertension → epistaxis
(iii) Angina (iv) Acute bronchitis
- Patient 5 (i) Chronic LVF (ii) Followed mitral incompetence due to mitral stenosis surgery* (iii) Cerebral embolus
(iv) Hypertension on treatment (v) Epileptic on treatment
- Patient 6 (i) Myocardial infarct (ii) Ventricular tachycardia*
- Patient 7 (i) Asthma on steroids precipitated by infection
(ii) Duodenal ulcer (iii) Haematemesis from steroids
(iv) Gastric surgery (v) Glaucoma (vi) Prostatism
- Patient 8 (i) Benign prostatic hypertrophy (ii) Hypochondriac
(iii) Depression
- Patient 9 (i) Subacute bacterial endocarditis
- Patient 10 (i) Nocturnal panic attacks (ii) Neurofibromatosis
(iii) Obsessive/depressive
- Patient 11 (i) Thyrotoxicosis (ii) L.orchidectomy (? tumour
? hydrocoele) (iii) Ca bronchus*

- Patient 12 (i) Chronic pulmonary fibrosis* (ii) Chronic bronchitis (iii) Emphysema
- Patient 13 (i) Myxoedema (ii) Reflux oesophagitis (iii) Familial optic atrophy
- Patient 14 (i) Nocturnal panic attacks (ii) Neurofibromatosis (iii) Obsessive/depressive
- Patient 15 (i) Myocardial infarct (ii) Diabetes (iii) Dermatitis (eczema)
- Patient 16 (i) Chronic rheumatic and ischaemic heart disease (ii) Paget's disease (iii) Penicillin allergy
- Patient 17 (i) Ca bronchus* (ii) Superior vena cava obstruction (iii) Recurrent laryngeal nerve involvement (iv) Prostatism
- Patient 18 (i) Drug-induced allergic jaundice
- Patient 19 (i) Stokes Adams attacks due to ischaemic heart disease
- Patient 20 (i) Recurrent myocardial infarct (ii) Dysrhythmia complicating
- Patient 21 (i) Syncopal episodes ? myocardial (ii) Cervical spondylosis (iii) Prostatism (iv) Urinary tract infection* (v) Treated hypertensive (vi) Acute bronchitis
- Patient 22 (i) Rheumatic heart disease (ii) Acute pulmonary oedema

NB * indicates that this diagnosis could not be made on the history alone in this patient.

APPENDIX 22

Diagnoses Given in each Patient's Medical Record - Cases Seen by House Officers

- Patient 1 (i) Hiatus hernia with reflux oesophagitis and stricture (ii) Chronic bronchitis with asthma
- Patient 2 (i) Primary biliary cirrhosis (ii) Scleroderma
- Patient 3 (i) Familial polycystic liver and pancreas with hepatic failure (ii) Steathorroea
- Patient 4 (i) Generalised atheromatous vascular disease (ii) Myocardial infarct (iii) Sinoatrial disease due to ischaemic heart disease (iv) Diabetes
- Patient 5 (i) Albright's syndrome* (ii) Osteosarcoma of femur* (iii) Depression
- Patient 6 (i) Chronic rheumatic heart disease (ii) Mitral valve replacement (iii) Past subacute bacterial endocarditis (iv) Stress incontinence
- Patient 7 (i) Diabetes (ii) Depression (iii) Arterio-pathic foot ulcer (iv) Cholecystitis
- Patient 8 (i) Haematemesis ? peptic ulcer (ii) Multiple nutritional deficiencies in a vegetarian (iii) Viral throat infection (iv) Malabsorption syndrome
- Patient 9 (i) Ca bronchus* (ii) Superior vena cava obstruction (iii) Recurrent laryngeal nerve involvement (iv) Prostatism
- Patient 10 (i) Chronic bronchitis with airways obstruction (ii) Bronchiectasis (iii) Respiratory failure (iv) Congestive cardiac failure

- Patient 11 (i) Angina (ii) Myocardial infarct
- Patient 12 (i) Chronic lymphatic leukaemia* (ii) Anxiety
- Patient 13 (i) Recurrent urinary tract infection with calculi (ii) Mixed pernicious/dietary anaemia (iii) Giving LVF (iv) Depression in social isolate
- Patient 14 (i) Chronic rheumatic heart disease (ii) Congestive cardiac failure (iii) ? Subacute bacterial endocarditis
- Patient 15 (i) Albright's syndrome* (ii) Osteosarcoma of femur* (iii) Depression (iv) Minor CVA → dysphasia
- Patient 16 (i) Treated duodenal ulcer (ii) Dumping syndrome (iii) Coeliac disease (iv) Dermatitis herpetiformis (v) Chronic bronchitis (vi) Prostatism
- Patient 17 (i) Sarcoidosis* (ii) Erythema nodosum
- Patient 18 (i) Myocardial infarction
- Patient 19 (i) Ca broncus* (ii) Cervical spondylosis
- Patient 20 (i) Ca bladder
- Patient 21 (i) Crohn's disease (ii) Subacute intestinal obstruction
- Patient 22 (i) Bronchopneumonia after falling and rib fracture

NB * indicates that this diagnosis could not be made on the history alone in this patient.

APPENDIX 23

Diagnoses Given in each Patient's Medical Record - Cases Seen by Registrars

- Patient 1 (i) Chronic myeloid leukaemia*
- Patient 2 (i) Chronic bronchitis with asthma (ii) Chronic duodenal ulcer (iii) Ischaemic heart disease
(iv) → recurrent supraventricular tachycardia
(v) → angina
- Patient 3 (i) Thyrotoxicosis (ii) Pernicious anaemia*
(iii) Cardiac failure from (i) and (ii)
- Patient 4 (i) Thyrotoxicosis (ii) Allergy to ATD
- Patient 5 (i) Chronic pancreatitis complicated by diabetes mellitus (ii) Myxoedema on treatment (iii) Blind loop syndrome following bypass operation for mistaken Ca pancreas*
- Patient 6 (i) Chronic myeloid leukaemia*
- Patient 7 (i) Myocardial infarct (ii) LVF
- Patient 8 (i) Intrinsic asthma (ii) Acute bronchitis
(iii) Allergic aspergillosis*
- Patient 9 (i) Crohn's disease
- Patient 10 (i) Acute pancreatitis → hepato renal failure
(ii) Secondary diabetes mellitus (iii) Alcoholic*
- Patient 11 (i) Crohn's disease
- Patient 12 (i) Pre-eclampsia of first pregnancy (ii) Ante-partum haemorrhage (iii) Acute tubular necrosis
(iv) Diffuse intravascular coagulation*

- Patient 13 (i) Osteomyelitis → leg amputation (ii) C.C.F.
(iii) Shingles (iv) Asymptomatic chronic duodenal
ulcer* (v) Haematemesis
- Patient 14 (i) Pulmonary tuberculosis
- Patient 15 (i) Duodenal ulcer (ii) Haematemesis (iii) Past
myocardial infarct
- Patient 16 (i) Past Ca breast (ii) Steroid-induced diabetes
(iii) Steroid-induced hypertension (iv) Depression
(v) Simulating paraplegia*
- Patient 17 (i) Chronic rheumatic heart disease (ii) Mitral
valve replacement (iii) Past subacute bacterial
endocarditis (iv) Stress incontinence (v) Post
operative cerebral embolus
- Patient 18 (i) Malabsorption (ii) Steatorrhoea
- Patient 19 (i) Myocardial infarct (ii) Reflux oesophagitis
(iii) Cervical spondylosis (iv) Depression
- Patient 20 (i) Duodenal ulcer (ii) Angina (iii) Cardiac
arrest from ischaemic heart disease (iv) Osteo-
arthritis of hip (v) Treated gout
- Patient 21 (i) Malnutrition (ii) Chronic bronchitis
(iii) Duodenal ulcer (iv) Broken ribs after
fall with exacerbation of chest infection
- Patient 22 (i) Pulmonary T.B. (ii) Malnutrition (iii) Epilepsy

NB *indicates that this diagnosis could not be made on the history
alone in this patient.

APPENDIX 24

Statistical Analyses of Patients' Diagnoses and Diagnoses Made Across Groups of Subjects

Table 1. Number of Diagnoses Given in Patients' Medical Records: Total, Range, Mean for Each Group (22 subjects per group)

	Total	Range	Mean	S.D.
Students	66	1-6	3	1.41
House Officers	63	1-6	2.86	1.32
Registrars	66	1-5	3	1.48

Table 2. Analysis of Variance Summary Table, Comparing Number of Patients' Diagnoses per Group

Source of variation	SS	Df.	Variance estimate
Between	0.26	2	0.13
Within	124.6	63	1.98
Total	124.86	65	F=0.06 (df. 2,63)NS.

Table 3. Numbers of Possible Diagnoses and Diagnoses Actually Made for Each Group of Subjects (raw data)

	DIAGNOSES GIVEN				DIAGNOSES MADE			
	Total	Range	Mean	SD	Total	Range	Mean	SD
Students	60	1-6	2.82	1.33	38	0-6	1.73	1.45
House Officers	55	1-6	2.54	1.40	25	0-3	1.23	1.02
Registrars	57	0-5	2.63	1.40	41	0-4	1.86	1.24

Table 4. Analysis of Variance Summary Table for the
Proportion of Correct Diagnoses Achieved by Each Subject

Source of variation	SS	Df.	Variance estimate
Between	7510	2	3755
Within	83993	63	1333
Total	91503	65	F=2.8169 (df. 2,63) NS.

APPENDIX 25

Preliminary Considerations in the Interpretation of Protocols

De Groot (1965) discusses the problem of protocol interpretation as follows:

"If theory or hypothesis testing is the research goal, ad hoc interpretations are taboo: predictions on introspective phenomena, like predictions on any other kind of phenomena, presuppose objective operationalizations. The procedures for protocol coding, in particular, must be objectively established in advance" (p. 383).

Although his subsequent discussion concerns only theory formation, not hypothesis testing, de Groot draws attention to a number of important points. For example, he mentions the need to avoid "arbitrary assumptions and decisions"; he asks us to consider "whether the effort required for meticulous protocol categorisation and analysis is well spent". He answers clearly in the affirmative. He also reminds us to be aware of the limits of justifiable interpretation in account analysis.

Elstein et al (1978) concentrate on considerations of a different order. They describe the four principles which directed the formulation of their scoring system (objectivity and reliability; task relevance; theoretical relevance; discriminant validity) and their method of developing content for that system:

"Our strategy was to develop the fundamental units of analysis from observations of physicians' performance in the setting of high-fidelity simulation, utilizing the episodic-review and stimulated-recall techniques to identify the fundamental elements in terms of which the solution of medical problems was organised. In the analysis of simulation data, much time was devoted to refining and sharpening the concepts by developing more operational definitions with satisfactory interrater reliability" (p. 51)

For some aspects of their protocols, nearly a year was spent without success in attempting to develop a scoring scheme. However, the

derivation of these was based on the protocols themselves and not on research hypotheses. This former approach is certainly more arduous than the latter and is also a necessary prerequisite stage (de Groot, 1965).

Richardson (1977), working in the entirely different field of housing research, faced the same problem of trying "to establish a methodology for dealing with data with no numbers attached". His decision "to look at people's talk about their experiences of changing environments as a primary source of data" caused him to crystallise the present problem:

"This decision, while being a solution to one set of research problems, was mother to a host of others, particularly in the area of numerical or statistical analysis. In abandoning structured instruments and questionnaires, we had also left behind the cosy bed of statistical package programmes and ease of translation of raw data to a form suitable for analysis".

Method of 'translation of raw data' is thus the first problem for solution, to be undertaken always in the light of the statistical options available for analysis of the translation. This method will invariably involve the development of a form of content categorisation system. Turney and Robb (1971) emphasise the importance of this process:

"The classification of data into convenient and logical categories is most critical in successfully reporting descriptive research. In categorising data, the researcher may well ask himself: What am I trying to describe? What are the elements that will contribute most to the description? What tables could best illustrate the description? Descriptive data can be collected in large quantities, and unless they are systematically dealt with, conclusions are difficult to draw. Ideally, one should be able to reduce data to quantitative terms and report them in tabular form. Where this is impossible, precise summaries are necessary".

Appendix 26 describes the method and process of content analysis applied to the accounts gathered for this study in an attempt to achieve Turney and Robb's ideal.

APPENDIX 26

The Method of Content Analysis

Holsti (1968) defines content analysis in the broadest terms as: "any technique for making inferences by systematically and objectively identifying specified characteristics of messages". The inferences to be made concern the antecedents of the communication; in this instance, the thinking process being explained by the subject. The validity of such an inferential process has been discussed and established (section 9.2), we therefore need only to describe the methodology.

The first decision to be made concerns the qualitative or quantitative use of content analysis. In essence we may say that the quantitative use of the method involves identifying frequency of occurrence of communication of any defined category, whereas qualitative use of the method merely requires identification of presence or absence of the defined category. Given our criticisms of others' use of accounts in terms of spurious quantification and unwarranted overinterpretation of data (see Chapter Three) and our awareness of the uncertainty of completeness and chronological accuracy of reports, it was considered unjustifiable to use a quantitative content analysis for each subject with one special exception (pre-diagnostic and diagnostic interpretations of clinical information). Our adopted approach, then, was primarily qualitative in that frequencies of any given category were not usually counted for each subject, but evidence was sought that the thinking process or content identified by any one category was either present or absent in any one subject. However, given the unusually large sample sizes, identification of presence or absence alone, which in one sense is the special case of quantification where the quantities are limited to one or zero, admits of statistical comparison of groups.

The next stage of development of the content analysis involves identifying categories of analysis for each hypothesis and indicators for the presence of each category.

The Formulation of Categories of Analysis and Identification of Indicators of Categories

Berelson (1971) points out that "since the categories contain the substance of the investigation, a content analysis can be no better than its system of categories". The development of a worthwhile set of categories is entirely dependent upon the presence of fully stated, clear and specific research hypotheses. This point seems to be emphasised above all others by writers on the subject:

"...relatively few ideas are "discovered" in the actual process of analysis. The hit-or-miss method of analysing "everything" in a body of content in the hope that "something will turn up" is seldom productive, and is certainly uneconomic. If the problem is not clarified to the point where several worthwhile hypotheses or questions can be formulated, then the projected analysis should be "abandoned" (p. 162).

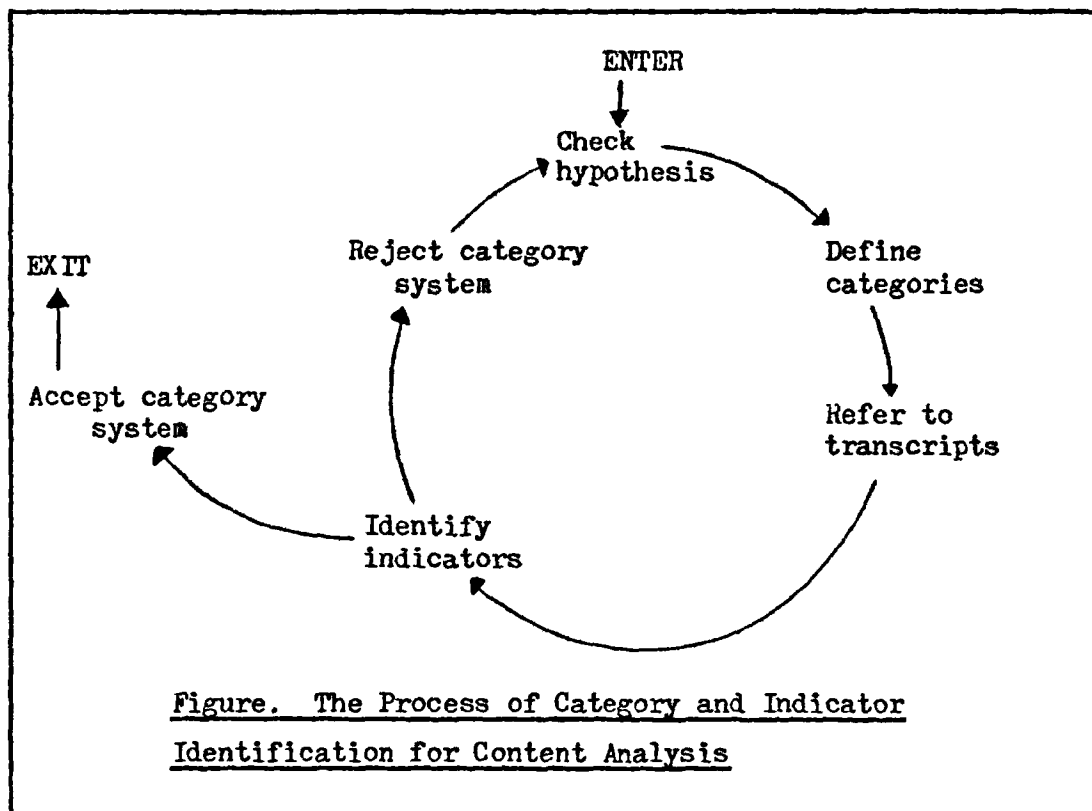
Carney (1972) also makes the point that content analysis cannot be used to "go fishing". Therefore, by a circular process of going from the research hypotheses to the transcripts and back again, a number of possible categories are identified for each hypothesis. Simultaneously, indicators for each category are listed. An indicator is merely a statement of the forms of words, with the content of those words if necessary, which would allow confirmation of presence of the associated category of thinking or response.

The categories themselves must have certain characteristics:

- i) They must suit both the hypotheses and the subject matter.
- ii) They must "set out clearly what sort of thing has to go into each of them and how that 'thing' is to be recognised" (Carney, 1972).

- iii) They must be inclusive enough to hold all appropriate items.
- iv) They must, between them, cover the whole range of issues pertinent to the enquiry.
- v) They have to be such that an item can be classified under only one of them for any one hypothesis.

Both Berelson (1971) and Carney (1972) make it quite clear that identification of a workable set of categories is very much a matter of trial and error. The process as it occurred in this study is summarised in the Figure. In all, this cycle on average for each hypothesis, occurred five times before a satisfactory set of categories and indicators emerged. With each successive refinement of categories and indicators the array becomes more workable and a greater number of transcripts (at one point, 63) must be read in order to disprove the system. Only when all 66 transcripts were read and categorised without problem was an array accepted as workable. It thus can appear to the researcher that the major feature of content analysis is its time consuming and circular developmental nature.



The process eventually yielded 14 separate categories for the seven hypotheses amenable to content analysis (Hypotheses 10 to 16). All research hypotheses are stated in section 5.1. However, only Hypotheses 10 to 20 refer to the account gathering study. Of these, Hypotheses 17 to 20 are framed in null terms as a matter of formality only. In essence, Hypotheses 17 to 20 ask whether or not the data furnish us with any evidence of certain types or features of cognitive process. For these hypotheses, therefore, content analysis in its present form was not applied. Instead, the transcripts were simply scanned in detail for possible evidence of such features as Hypotheses 17 to 20 identify. Content analysis in the form described was not appropriate since the research hypotheses give no indication of possible categories. Hypotheses 17 to 20, therefore, demand a process which de Groot (1965) refers to as 'theory formation' rather than 'hypothesis testing'. Hypotheses 10 to 16, however, reflect this more advanced research stage of hypothesis testing and so require the process of content analysis.

The content of the 14 categories is presented as results (section 10.1).

APPENDIX 27

Establishment of Reliability and Validity of Content Analysis Categories and Raters

Each content analysis was performed by one person only. When technical medical expertise was required in the identification of pre-diagnostic and diagnostic interpretations, the rater was a consultant physician. All other analyses were performed by the author. In both cases reliability of raters was established by reference to other judges and test-retest reliability was established by repeated ratings. In addition, the classification of subjects' interpretations into pre-diagnostic and diagnostic categories was validated by four other consultant physician raters. Each of these tests may be considered in turn.

Validity and Reliability of Pre-diagnostic and Diagnostic Interpretations Categories (Categories (a) and (b) for Hypotheses 10, 11 and 12)

The experimenter and primary rater independently identified all interpretations of clinical information made by each subject. There was 100 per cent agreement between those two raters on this. The interpretations were ordered alphabetically for each group of subjects. Students made 175 different interpretations, house officers 200 and registrars 235. A total of 610 interpretations was therefore listed. Each interpretation was classified by the primary rater as either 'pre-diagnostic' or 'diagnostic' according the definitions of those terms given (see section 10.1.1). One in six (approximately 17 per cent) of these interpretations was taken from all the alphabetical lists and these 105 formed into an inter-rater and test-retest reliability form (see Appendix 28 for the items thus selected and the designated category of each item). The representativeness of the sample of items is shown in Table 1. Approximately the same ratio of pre-diagnostic to diagnostic interpretations occur in both population and sample.

Table 1. Characteristics of Population and Sample of Pre-diagnostic and Diagnostic Interpretations.

	<u>Pre-diagnostic</u>	<u>Diagnostic</u>	<u>Total</u>
Population	387 (63%)	223 (37%)	610
Sample	64 (61%)	41 (39%)	105

The form was distributed to the primary rater and four consultant physicians, with definitions of the categories. Each respondent assigned each items to one category. The primary rater's test-retest reliability obtained by repeated ratings was 100 per cent. Results of the inter-rater reliability test are shown in Table 2.

Table 2. Results of Inter-rater Reliability Study of Pre-diagnostic and Diagnostic Categories

	4 out of 4 raters (100 per cent agreement)	3 out of 4 raters (75 per cent agreement)	2 out of 4 raters (50 per cent agreement)
Number of items	87	14	4
Percent of items	83	13	4

Total number of items rated - 420

Total number of disagreed ratings - 64

Overall agreement - 85 per cent

Overall inter-rater reliability is seen to be 85 per cent. Although one may have hoped for a higher value than this, in the circumstances this would probably be unrealistic. For example, a number of disagreements arose when the items under consideration impinged on the

area in which a rater was a specialist (for example: "thyrotoxicosis", "coronary") and whereas others would accept these as 'diagnostic', the specialist would not. It also occurred that the inaccuracy and incorrectness of some items offended and provoked replies of 'impossible' only (for example: 'LVF secondary to mitral stenosis', 'cardiac problems secondary to asthma'). Given these responses, an inter-rater reliability of 85 per cent is accepted as satisfactory.

Inter-rater and Repeat Reliability of all other Categories

These forms of reliability were established using a sample of seven transcripts (11 per cent sample). Two transcripts were of registrars, two of house officers and three of students. Each transcript was subjected to content analysis using the remaining 12 categories by both the experimenter (repeat reliability) and an independent rater. Repeat reliability was 100 per cent. Table 3 shows the results of inter-rater reliability for each category.

<u>Table 3. Results of Inter-rater Reliability Study for the Remaining 12 Categories</u>							
Category	11(c)	12(c)	13(a)	13(b)	14(a)	14(b)	14(c)
Number agreed (out of 7)	7	7	6	7	6	7	6
Category	15(a)	15(b)	15(c)	16(a)	16(b)		
Number agreed (out of 7)	7	7	7	7	7		
Total number of categories rated				-	84		
Total number of disagreed ratings				-	3		
Overall agreement				-	96 per cent		

An overall reliability of 96 per cent is shown, with 100 per cent agreement on nine categories, and six out of seven agreement on the remaining three categories. Such a level of agreement is accepted as satisfactory and indicative of both validity of the categories and reliability of their use.

Validity and reliability of the content analysis category system is thus established. Given the demonstrated validity of case content across groups (section 9.2.3 and Appendix 24) and subjects' estimation of the validity of their own performances (Table 9.5), all necessary statistical parameters have been applied to the method and encourage confidence in both its validity and reliability of implementation in this study.

APPENDIX 28

Items with Designated Categories Selected for Reliability Study
of Pre-diagnostic (P-d) and Diagnostic (D) Interpretations

<u>ITEM</u>	<u>CATEGORY</u>
Weak ankle	P-d
Anaemia due to uterine bleeding	D
Arrhythmia	P-d
Mesenteric angina	D
Antepartum haemorrhage	D
Biliary colic	D
Chronic bronchitis	D
Gastric carcinoma	D
Congestive cardiac failure	P-d
Coronary	D
Ischaemic colitis	D
Carcinoma of the breast with secondaries in the back	D
Antacid induced diarrhoea	D
Dysentery	P-d
An emotional thing	P-d
Gall bladder problems	P-d
Prednisolone induced hypertension	D
Hiatus hernia	D
Haemochromatosis	D
Iron overdose	D
Paralytic ileus	P-d
Jod-Basedow phenomenon	D
Lassitude	P-d
Chronic lymphatic leukaemia	D
Lumbar puncture	P-d
Myaesthesia from carcinoma	P-d
Myxoedema madness	D
Myocardial infarct	D
Nephrotic syndrome	P-d
Something neurological	P-d

<u>ITEM</u>	<u>CATEGORY</u>
An organic thing	P-d
Pyloric stenosis secondary to carcinoma of the stomach	D
Pneumonic elements	P-d
Peritonitis	P-d
Psychological problems	P-d
Respiratory disease	P-d
Root pain	P-d
Renal biopsy	P-d
Something wrong with the stomach	P-d
Retinal vein thrombosis	D
Gastric ulcer	D
Varicose ulcer	D
Weakness from lying in bed	P-d
Anaemia caused by blood loss	P-d
Anaemia secondary to malignancy	P-d
Chronic obstructive airways disease	D
Asbestosis	D
Blood pressure	P-d
Apical lung carcinoma	D
Chest trouble	P-d
Calcium metabolism defect	P-d
Cardiac failure	P-d
CVA	P-d
Some reason for cardiac drugs	P-d
Diverticular disease	D
Familial bone disease	P-d
Gutty thing	P-d
Hypothermic	P-d
On iron tablets	P-d
Iritis	P-d
Liver disease	P-d
LVF secondary to mitral stenosis	D
Malnutrition	P-d
Autonomic nervous system giddy spells	P-d

<u>ITEM</u>	<u>CATEGORY</u>
Oedema	P-d
Psoriasis	D
Prostatism	P-d
Polyuria/polydipsia related to calcium	P-d
Bone pain	P-d
Rheumatoid	P-d
Syncope	P-d
TB	D
Underlying constitutional illness	P-d
Peripheral vascular disease	P-d
Aortic valve disease	P-d
Atheroma	P-d
Asthma	D
Steroid induced Cushing's	D
Cardiac problem secondary to asthma	P-d
Carcinomatosis	P-d
Paroxysmal dyspnoea	P-d
Dysphasia	P-d
Ear trouble	P-d
Grand Mal epilepsy	D
A fit	P-d
Hypoglycaemia	P-d
Hay fever	D
Indigestion	P-d
Lung abscess	D
Meningitis complicating labarynthitis	D
Myxoedema	D
Neurofibroma	P-d
Oedema from subcutaneous intravenous fluid	D
Porphyria	D
Respiratory tract infection	P-d
Gastric reflux	P-d
Bony secondaries	P-d
Thrombotic episode	P-d
Acute hepatitis	D

<u>ITEM</u>	<u>CATEGORY</u>
Haematemesis from peptic ulcer	D
Raised intracranial pressure headache	P-d
Thyrotoxicosis	D
Uraemia	P-d
Virus infection	P-d
Valvular defect	P-d

APPENDIX 29

Content Analysis Raw Data: Observed Frequencies per Category
per Group

Hypothesis	Category	Frequency per Group (n=22)			TOTAL n=66
		Students	House Officers	Registrars	
10	a	154	164	165	483
	b	91	96	114	301
11	a	10	17	15	42
	b	7	8	7	22
	c	9	4	6	19
12	a	22	22	21	65
	b	21	22	22	65
	c	22	22	22	66
13	a	3	6	4	13
	b	12	14	12	38
14	a	18	11	12	41
	b	14	14	13	41
	c	12	13	15	40
15	a	6	1	1	8
	b	21	20	21	62
	c	21	21	18	60
16	a	12	8	4	24
	b	3	1	4	8

APPENDIX 30

Raw Data of Content Analysis using Categories 10(a) (pre-diagnostic interpretations) and 10(b) (diagnostic interpretations), Recording the Exact Number of Each Type per Subject per Group

Subject Number	Students			House Officers			Registrars		
	10(a)	10(b)	Total	10(a)	10(b)	Total	10(a)	10(b)	Total
01	7	0	7	2	1	3	10	3	13
02	10	4	14	7	4	11	17	7	24
03	2	5	7	10	1	11	7	7	14
04	10	3	13	17	1	18	7	4	11
05	18	1	19	4	3	7	3	4	7
06	9	4	13	3	4	7	18	10	28
07	6	8	14	8	0	8	2	5	7
08	17	4	21	7	7	14	4	2	6
09	8	2	10	12	3	15	7	3	10
10	7	4	11	13	10	23	5	5	10
11	2	9	11	8	5	13	3	6	9
12	6	3	9	9	6	15	16	9	25
13	4	4	8	12	4	16	6	8	14
14	8	4	12	5	8	13	7	5	12
15	1	3	4	9	3	12	7	4	11
16	4	5	9	6	7	13	7	6	13
17	10	7	17	10	7	17	0	3	3
18	2	5	7	6	4	10	5	6	11
19	2	5	7	7	11	18	9	7	16
20	6	3	9	2	1	3	8	6	14
21	11	3	14	4	1	5	8	3	11
22	4	5	9	3	5	8	9	1	10
Total	154	91	245	164	96	260	165	114	279
Mean	7	4.1	11.1	7.4	4.4	11.8	7.5	5.2	12.7
S D	4.56	2.05	4.22	3.87	3.05	5.15	4.57	2.28	6.10

REFERENCES

- ABERCROMBIE, M.L.J. (1960) The Anatomy of Judgement. Harmondsworth: Penguin.
- ABERCROMBIE, M.L.J. (1978) Improving teaching and learning in small groups. Med. Educ., 12, 5, Supplement, 68-69.
- ABERCROMBIE, M.L.J., COLES, C., GALE, JANET, WAKEFORD, R.E. and WYN PUGH, E. (1978) Three approaches to curriculum evaluation in British Medical Schools. Paper presented at the September ASME Scientific Meeting, Royal Free Hospital, London.
- ACHESON, E.D. (1976) Southampton: the first years. 1 - About Southampton Medical School. Brit. Med. J., 2, 23-25.
- AITCHISON, J. (1970) Decision-making in clinical medicine. J. of the Roy. Coll. of Physicians of London, 4, 3, 195-202.
- AITCHISON, J. and KAY, J.W. (1973) Principles, practice and performance in decision-making in clinical medicine. Paper prepared for the NATO Conference on 'The Role and Effectiveness of Decision Theories in Practice', Luxembourg.
- ALBERT, D.A. (1974) The logical structure of the patient history. J. of Med. Educ., 49, 12, 1152-1157.
- ANDERSON, N.H. and SHANTEAU, J.C. (1970) Information integration in risky decision making. J. of Exp. Psychol., 84, 3, 441-451.
- ARMSTRONG, D. (1977) The structure of medical education. Med. Educ., 11, 4, 244-248.
- ARMSTRONG, D., HICKS, B.H., HIGGINS, P.M. and WEINMAN, J.A. (1979) Teaching communication skills to pre-clinical medical students: A general practice based approach. Med. Educ., 13, 2, 82-85.
- ASSOCIATION FOR THE STUDY OF MEDICAL EDUCATION (1973) Curriculum Changes in United Kingdom Medical Schools. Conference papers. Dundee: ASME.
- AUSUBEL, D.P., NOVAK, J.D. and HANESIAN, H. (1978) Educational Psychology. A Cognitive View. New York: Holt, Rinehart and Winston.
- BARROWS, H.S. and BENNETT, K. (1972) The diagnostic (problem solving) skill of the neurologist. Arch. of Neurol., 26, 273-277.
- BARROWS, H.S. (1973) Problem-based learning in medicine: Rationale and methods. McMaster University Education Monograph 4
- BARROWS, H.S. and MITCHELL, D.L.M. (1975) An innovative course in undergraduate neuroscience: Experiment in problem-based learning with 'problem boxes'. Brit. J. of Med. Educ., 9, 4, 223-230.
- BARROWS, H.S. (1976) Keynote address: Problem-based learning in medicine. In J. Clarke and J. Leedham (eds) Aspects of Educational Technology 10: Individualised Learning. London: Kogan Page.
- BARROWS, H.S. and TAMBLYN, R.M. (1977) The portable patient problem pack: A problem-based learning unit. J. of Med. Educ., 52, 1002-1004.
- BARROWS, H.S., NORMAN, G.R., NEUFELD, V.R. and FEIGHTNER, J.V. (1978) The clinical reasoning process of the physician. Unpublished report.
- BART, J.M. and SMITH, M.B. (1974) An interpretive framework of cognitive structures. Human Dev., 17, 3, 161-175.

- BARTLETT, F.C. (1958) Thinking. An Experimental and Social Study. London: George Allen and Unwin Ltd.
- BASHOOK, P.G. (1976) A conceptual framework for measuring clinical problem-solving. J. of Med. Educ., 51, 2, 109-114.
- BERELSON, B. (1971) Content Analysis in Communication Research. New York: Hafner Publishing Co.
- BERNER, E.S. and BLIGH, T.J. (1974) How content specific is medical problem solving? Report to the Faculty, Center for Educational Development. University of Illinois College of Medicine.
- BERNER, E.S., HAMILTON, L.A. and BEST, W.R. (1974) A new approach to evaluating problem-solving in medical students. J. of Med. Educ., 49, 666-672.
- BERNER, E.S., BLIGH, T.J. and GUERIN, R.O. (1977) An indication for a process dimension in medical problem-solving. Med. Educ., 11, 5, 324-328.
- BERNSTEIN, B. (1971) On the classification and framing of educational knowledge. In M.F.D. Young (Ed) Knowledge and Control. London: Collier MacMillan.
- BIRCH, H.G. and RABINOWITZ, H.S. (1951) The negative effect of previous experience on productive thinking. In P.C. Wason and P.N. Johnson-Laird (Eds) (1968) Thinking and Reasoning. Harmondsworth: Penguin Books Ltd.
- BLOOM, B.S. (1953) Thought processes in lectures and discussions. J. of Gen. Educ., 7, 160-169.
- BLOOM, B.S. (1956) Taxonomy of Educational Objectives. New York: Longmans.
- BREHMER, B. (1973) Note on clinical judgement and the formal characteristics of clinical tasks. Umeå Psychological Reports, No. 77, University of Umeå, Sweden.
- BRITISH MEDICAL JOURNAL (1977) Reducing doctors' errors. Editorial Brit. Med. J., 1, 1178-1179.
- BROADBENT, D.E. (1975) Cognitive psychology and education. Brit. J. of Educ. Psychol., 45, 2, 162-176.
- BRUNER, J.S. (1951) Personality dynamics and the process of perceiving. In R.R. Blake and G.V. Ramsey (Eds) Perception - An Approach to Personality. New York: The Ronald Press Company.
- BRUNER, J.S., GOODNOW, J.J. and AUSTIN, G.A. (1956) A Study of Thinking. New York: John Wiley and Sons Inc.
- BRUNER, J.S. (1957) Going beyond the information given. In Contemporary Approaches to Cognition. A Symposium held at the University of Colorado. Cambridge, Mass: Harvard University Press.
- BRUNER, J.S. (1960) The Process of Education. Cambridge, Mass: Harvard University Press.
- BRUNER, J.S. (1964) Some theorems on instruction illustrated with reference to mathematics. In E.R. Hilgard (Ed) Theories of Learning and Instruction. Chicago, Illinois: University of Chicago Press.
- BRUNER, J.S. (1966) Toward a Theory of Instruction. Cambridge, Mass: The Belknap Press.

- BRUNER, J.S. (1973) Beyond the Information Given. London: George Allen and Unwin Ltd.
- BUCKLEY-SHARP, M.D. and HARRIS, F.T.C. (1971) The scoring of multiple choice questions. Brit. J. of Med. Educ., 5, 4, 279-288.
- BUTCHER, H.J. (1968) Human Intelligence: Its Nature and Assessment. London: Methuen and Co. Ltd.
- BYRNE, P.S. and FREEMAN, J. (1971) Clinical factual recall in general practice. Brit. J. of Med. Educ., 5, 2, 118-124.
- CAMPBELL, D.T. and STANLEY, J.C. (1963) Experimental and Quasi-experimental Designs for Research. Chicago: Rand McNally and Co.
- CARD, W.I. (1970) The diagnostic process. J. of the Roy. Coll. of Physicians of London, 4, 3, 183-187.
- CARNEY, T.F. (1972) Content Analysis: A Technique for Systematic Inference from Communications. London: B.T. Batsford Ltd.
- CASE, R. (1973) Piaget's theory of child development and its implications. Phi Delta Kappan, September, 20-25.
- CLARKE, R., FENTEM, P. and SHORT, T. (1976) Multiple choice questions (Letter). Med. Educ., 10, 3, 243.
- CLARKE, R.M. (1978) Curricular innovation at the new medical school at Newcastle, Australia. Med. Educ., 12, 5, Supplement, 53-56.
- CLOUSER, K.D. (1977) Clinical medicine as science. Editorial. J. of Med. and Philosophy, 2, 1, 1-7.
- COHEN, J. (1972) Psychological Probability and the Art of Doubt. London: George Allen and Unwin Ltd.
- CONNOLLY, J. and BIRD, J. (1977) Videotape in teaching and examining clinical skills: A short case format. Med. Educ., 11, 4, 271-275.
- COOLEY, W.W. and LOHNES, P.R. (1971) Multivariate Data Analysis. New York: John Wiley and Sons Inc.
- CORCORAN, D.W.J. (1971) Pattern Recognition. Harmondsworth: Penguin Books Ltd.
- CREHAN, K.D., CANDOR, C.A. and BECKETT, G.W. (1976) Utilisation of partial knowledge of objective exams under formula - and number-right scoring directions. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- CRONBACH, L.J. (1960) Essentials of Psychological Testing. New York: Harper Row.
- CROPPER, D.A., MECK, D.S. and ASH, M.J. (1977) The relation between formal operations and a possible fifth stage of cognitive development. Dev. Psychol., 13, 5, 517-518.
- CROSS, G.R. (1974) The Psychology of Learning. An Introduction for Students of Education. Oxford: Pergamon Press.
- CRO'S, L.H. and FRARY, R.B. (1977) An empirical test of Lord's theoretical results regarding formula scoring of multiple choice tests. J. of Educ. Meas., 14, 4, 313-321.

- CRUTCHFIELD, R.S. (1972) Nurturing the cognitive skills of productive thinking. In M.L. Silberman, J.S. Allender and J.M. Yanoff (Eds) The Psychology of Open Teaching and Learning. An Inquiry Approach. Boston: Little, Brown and Company.
- CUTLER, P. (1979) Problem Solving in Clinical Medicine. From Data to Diagnosis. Baltimore: The Williams and Wilkins Company.
- de DOMBAL, F.T., HARTLEY, J.R., and SLEEMAN, D.H. (1969) A computer-assisted system for learning clinical diagnosis. The Lancet, 1, 145-148.
- de DOMBAL, F.T., HORRICKS, J.C., STANILAND, J.R. and GUILLOU, P.J. (1971) Production of artificial 'case histories' by using a small computer. Brit. Med. J., 2, 578-581.
- de DOMBAL, F.T., LEAPER, D.J., STANILAND, J.R., McCANN, A.P. and HORROCKS, J.C. (1972a) Computer-aided diagnosis of acute abdominal pain. Brit. Med. J., 2, 9-13.
- de DOMBAL, F.T., SMITH, R.B., MODGILL, V.K. and LEAPER, D.J. (1972b) Simulation of the diagnostic process: A further comparison. Brit. J. of Med. Educ., 6, 3, 238-245.
- de DOMBAL, F.T., HORROCKS, J.C., CLAMP, S.F. and STORR, J.E. (1974) Simulation techniques and computer-based teaching of the clinical diagnostic process. Medinfo '74. Amsterdam: North Holland Publishing Co.
- de DOMBAL, F.T. (1979) Computers and the surgeon - a matter of decision. Surgery Annual, 11, 1, 33-57.
- de GROOT, A.D. (1965) Thought and Choice in Chess. The Hague: Mouton and Co.
- de GROOT, A.D. (1966) Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed) Problem Solving: Research, Method and Theory. New York: Wiley.
- LICKSTEIN, L.C. (1968) Field independence in concept attainment. Percept. and Motor Skills, 27, 2, 635-642.
- DIENES, Z.P. and JEEVES, M.A. (1965) Thinking in Structures. London: Hutchinson Educational.
- DONALDSON, M. (1959) Positive and negative information in matching problems. Brit. J. of Psychol., 50, 2, 235-262.
- DREVER, J. (1952) A Dictionary of Psychology. Harmondsworth: Penguin Books Ltd.
- DUDLEY, H.A.F. (1969) Tasks for clinicians: the diagnostic process. Med. J. of Australia, 1, 37-43.
- DUDLEY, H.A.F. (1970) The clinical task. The Lancet, 2, 1352-1354.
- DUDLEY, H.A.F. (1971) Clinical method. The Lancet, 1, 35-37.
- DUNCKER, K. (1945) On problem-solving. Psychol. Monogr., 58, 270.
- EK O, E.E. and LOFNING-BAUCKE, V. (1979) Clinical problem solving and clinical knowledge. Med. Educ., 13, 4, 251-256.
- ELLIS, J. (1960) Preparation for the profession. The Lancet, 2, 1040-1045.
- FLICKSTEIN, A.S., KAGAN, N., SHULMAN, L.S., JACON, H. and LOUPE, M. (1972) Methods and theory in the study of medical enquiry. J. of Med. Educ., 47, 85-92.

- ELSTEIN, A.S. (1976) Clinical judgement: Psychological research and medical practice. Science, 194, 4266, 696-700.
- ELSTEIN, A.S., SHULMAN, L.S., VINSONHALER, J., WAGNER, C. and BADER, L. (1977) Characteristics of the clinical problem solving model and its relevance to educational research. Paper presented at the Annual Meeting of the American Educational Research Association, New York.
- ELSTEIN, A.S., SHULMAN, L.S. and SPRAFKA, S.A. (1978) Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, Mass.: Harvard University Press.
- ELSTEIN, A.S., SPRAFKA, S.A. and BORDAGE, G. (1979) Problem solving: Application of research to undergraduate instruction and evaluation. Prog. Learning and Educ. Technol., 16, 4, 296-302.
- ELSTEIN, M. and FORBES, J.A. (1976) Southampton: the first years. 4 - Early medical contact. Brit. Med. J., 2, 97-98.
- ENGEL, C.E. and CLARKE, R.M. (1979) Medical education with a difference. Prog. Learning and Educ. Tech., 16, 1, 70-87.
- ENGLE, R.L. (1964) Medical diagnosis. In J.A. Jacquez (Ed) The Diagnostic Process. Proceedings of a conference sponsored by the Biochemical Data Processing Training Program of the University of Michigan. Ann Arbor, Michigan: Malloy Lithographing Inc.
- FISCHER, B.J. (1976) Diagnostic Pathways in Clinical Medicine. London: Churchill - Livingstone.
- EVANS, J. St. B.T. (1976) A critical note on Quinton and Fellows' observation of reasoning strategies. Brit. J. of Psychol., 67, 4, 517-518.
- EVANS, J. St. B.T. and JASON, P.C. (1976) Rationalisation in a reasoning task. Brit. J. of Psychol., 67, 4, 479-486.
- FARQUHAR, J.A., BRETZ, R., GINSBURG, A.S., LINCOLN, T.L., MELLONE, R.J., and MILLS, G.F. (1970) Applications of Advanced Technology to Undergraduate Medical Education. Memorandum RM - 6180 - NLM. Prepared for the National Library of Medicine. Santa Monica, California: The Rand Corporation.
- FEINSTEIN, A.R. (1963) Boolean algebra and clinical taxonomy 1. Analytic synthesis of the general spectrum of a human disease. The New Eng. J. of Med., 18, 929-938.
- FEINSTEIN, A.R. (1967) Clinical Judgement. Baltimore: Williams and Wilkins.
- FELLOWE, B.J. (1976) The role of introspection in problem-solving research: A reply to Evans. Brit. J. of Psychol., 67, 4, 519-520.
- FUSON, G.A. (1966) Statistical Analysis in Psychology and Education. London: McGraw-Hill.
- ITIN, P.M. and POENLER, M.I. (1967) Human Performance. Belmont, California: Brooks/Cole Publishing Company.
- MAHERTY, F.G. (1975) The thinking aloud technique and problem solving ability. J. of Educ. Res., 68, 6, 223-225.
- LAVERL, J.H. (1963) The Developmental Psychology of Jean Piaget. Princeton, New Jersey: D. Van Nostrand Company Inc.
- LEWIS, P.N., ANDERSON, P.H., STOKES, J., and ALTON, H.J. (1976) Examinations in Medicine. London: Churchill-Livingstone.

- FLEXNER, A. (1925) Medical Education. A Comparative Study. New York: MacMillan.
- FORSTROM, L.A., (1977) The scientific autonomy of clinical medicine. J. of Med. and Philosophy, 2, 1, 8-19.
- FREER, C.B. (1978) Videotape recording in the assessment of the history-taking skills of medical students. Med. Educ., 12, 5, 360-363.
- FURTH, H.G. (1968) The nature of representation and interiorization. Psychol. Rev., 75, 143-154.
- GAGE, N.L. (1963) Paradigms for research on teaching. In N.L. Gage (Ed) Handbook of Research on Teaching. Chicago: Rand McNally and Co.
- GAGE, N.L. (1964) Theories of teaching. In E.R. Hilgard (Ed) Theories of Learning and Instruction. Chicago, Illinois: University of Chicago Press.
- GAGNÉ, R.M. (1964) Problem solving. In A.H. Melton (Ed) Categories of Human Learning. New York: Academic Press.
- GAGNÉ, R.M., (1965) The Conditions of Learning. London: Holt, Rinehart and Winston.
- GAGNÉ, R.M. (1966) Human problem solving: Internal and external events. In B. Kleinmuntz (Ed) Problem Solving: Research, Method and Theory. New York: John Wiley and Sons Inc.
- GARLAND, L.H. (1959) Studies on the accuracy of diagnostic procedures. Amer. J. of Roentgen., 82, 25-38.
- GARNER, W.R. (1962) Uncertainty and Structure as Psychological Concepts. New York: John Wiley and Sons Inc.
- GENERAL MEDICAL COUNCIL (1967) Recommendations as to Basic Medical Education. London: GMC.
- GENERAL MEDICAL COUNCIL (1977) Basic Medical Education in the British Isles. Vols. 1 and 2. London: Nuffield Provincial Hospitals Trust.
- GETZELS, J.W. (1964) Creative thinking, problem-solving and instruction. In E.R. Hilgard (Ed) Theories of Learning and Instruction. Chicago, Illinois: University of Chicago Press.
- GOEDBERG, L.R. (1968) Simple models or simple processes? Some research on clinical judgements. Amer. Psychologist, 23, 483-496.
- GOEDBERG, L.R. (1970) Man versus model of man: A rationale plus some evidence for a method of improving on clinical inferences. Psychol. Bull., 73, 6, 422-432.
- GOULD, G.A. (1968) Sequential diagnosis by computer. J. of the Amer. Med. Assoc., 25, 12, 849-854.
- GOULD, G.A. (1970) Modelling the diagnostic process. J. of Med. Educ., 45, 293-302.
- GRONLUND, N.E. (1971) Measurement and Evaluation in Teaching. New York: MacMillan.
- GUILFORD, J.P. (1954) Psychometric Methods. New York: McGraw-Hill.
- HAGGARD, F.A. (1958) Intraclass Correlation and the Analysis of Variance. New York: Dryden Press Inc.
- HAMILTON, M. (1966) Clinicians and Decisions. Leeds: Leeds University Press.
- HAMMOND, K.R. and SUMMERS, D.A. (1965) Cognitive dependence on linear and non-linear cues. Psychol. Rev., 72, 215-224.
- HAMMOND, K.R. (1966) Clinical inference in nursing. 2 A psychologist's viewpoint. Nursing Res., 15, 1.

- HAIMOND, K.R. (1971) Computer graphics as an aid to learning. Science, 172, 903-908.
- HAIMOND, L.K. (1970) Cognitive structure and clinical inference. J. of Gen. Psychol., 83, 107-113.
- HAIMPTON, J.R., HARRISON, M.J.G., MITCHELL, J.R.A., PRITCHARD, J.S. and SEYMOUR, C. (1975) Relative contributions of history-taking, physical examination and laboratory investigation to diagnosis and management of medical outpatients. Brit. Med. J., 2, 486-489.
- HARDEN, R.McG., STEVENSON, H., DONNIE, J. and TILSON, G.M. (1975) Assessment of clinical competence using objective, structured examination. Brit. Med. J., 1, 447-451.
- HARDEN, R. McG., BROWN, R.A., BIRAN, L.A., DALLAS MOSS, W.P. and AKEFORD, R.E. (1976) Multiple choice questions: to guess or not to guess. Med. Educ., 10, 1, 27-32.
- HARDEN, R. McG., ROGER, J. and WAKEFORD, R.F. (1978) Medical education in the British Isles: trends identified in the GMC survey of basic medical education. Med. Educ., 12, 5, Supplement, 25-28.
- HARDEN, R. McG. and GLEESON, F.A. (1979) Assessment of Clinical Competence Using an Objective Structured Clinical Examination (OSCE). ASME Medical Education Booklet No. 8. Dundee: ASME.
- HARLESS, W.G., DRENNON, G.G., MARKER, J.H., ROOT, J.A. and MILLER, G.E. (1971) CASE: A computer-aided simulation of the clinical encounter. J. of Med. Educ., 46, 443-448.
- HARRIS, A., LAMN, M. and PRESCOTT, W. (1975) (Eds) Curriculum Innovation. London: Croom Helm and The Open University Press.
- HARRISON, T.R., VINTROBE, M.M., KAHLER, R.I. and BRAUNWALD, E. (1971) Cyanosis, hypoxia and polycythaemia. In M.M. Vintrobe, G.W. Thorn, R.E. Adams, I.L. Bennett, E. Braunwald, K.J. Isselbacher and R.G. Petersdorf (Eds) Harrison's Principles of Internal Medicine. New York: McGraw-Hill.
- HAYS, J.L. (1963) Statistics. London: Holt, Rinehart and Winston.
- HELPER, R.E. and SLATER, C.H. (1971) Measuring the process of solving clinical diagnostic problems. Brit. J. of Med. Educ., 5, 48-52.
- HILGARD, E.R. (1964) (Ed) Theories of Learning and Instruction. Chicago, Illinois: University of Chicago Press.
- HILL, G.C. and MOOLS, G.T. (1974) Multiple true-false questions. Educ. in Chem., 11, 86-87.
- HINZ, C.F. (1966) Direct observation as a means of teaching and evaluating clinical skills. J. of Med. Educ., 41, 2, 150-161.
- HISS, R.G. and PIERCE, J.C. (1974) A strategy for developing educational objectives in medicine: Problem-solving skills. J. of Med. Educ., 49, 660-665.
- HOFFMAN, P.J. (1960) The paramorphic representation of clinical judgement. Psychol. Bull., 57, 116-131.
- HOFFMAN, P.J. (1968) Cue consistency and configurality in human judgement. In B. Kleinmuntz (Ed). Formal Representation of Human Judgement. New York: Wiley.
- HOFFMAN, P.J., S (VIC, P. and OBER, T.G. (1968) An analysis of variance model for the assessment of configural cue utilization in clinical judgement. Psychol. Bull., 69, 3, 338-349.
- HUGHES, F.F., BAKER, L.H., TORIAN, E.C., RICHARDSON, N.K., GUCK, S. and YARAT, A.J. (1978) Measuring clinical competence of medical students. Med. Educ., 12, 5, 364-368.

- HOLSTI, O.R. (1968) Content Analysis. In G. Lindzey and E. Aronson (Eds) Handbook of Social Psychology Vol. 2 Research Methods. 2nd edition. Massachusetts: Addison-Wesley Publishing Company.
- HOLT, R.R. (1958) Clinical and statistical prediction: A reformulation and some new data. J. of Abnorm. and Soc. Psychol., 56, 1-12.
- HOLT, R.R. (1970) Yet another look at clinical and statistical prediction: Or is clinical psychology worthwhile? Amer. Psychologist, 25, 337-349.
- HOOPER, R. (1971) (Ed) The Curriculum: Context, Design and Development. Edinburgh: Oliver and Boyd and The Open University Press.
- HOPE, K. (1968) Methods of Multivariate Analysis. London: University of London Press Ltd.
- HOWE, M.J.A. and Colley, L. (1976) The influence of questions encountered earlier on learning from prose. Brit. J. of Educ. Psychol., 46, 2, 149-154.
- HOWELL, J.B.I. (1976) Southampton: the first years. 2 - Systems courses. Brit. Med. J., 2, 26-27.
- HUBBARD, J.P. and CLEMANS, W.V. (1961) Multiple-choice Examinations in Medicine. Philadelphia: Lea and Febiger.
- HUBBARD, J.P. (1971) Measuring Medical Education. The Tests and Test Procedures of the National Board of Medical Examiners. Philadelphia: Lea and Febiger.
- HUFF, F.W. and FRIEDMAN, H. (1967) A study of the effects on agreement among clinicians of redundant and new information, confidence and time available for assessments. J. of Gen. Psychol., 76, 49-57.
- HUNTER, I.M.L. (1964) Memory. Harmondsworth: Penguin Books Ltd.
- IANCEK, R., and BALIA, J.I. (1979) The neurological diagnostic process: We should preach what we practise. Unpublished paper.
- INHELDER, B. and PIAGET, J. (1958) The Growth of Logical Thinking from Childhood to Adolescence. London: Routledge and Kegan Paul.
- JACQUEZ, J.A. (Ed) The Diagnostic Process. Proceedings of a conference sponsored by the Biochemical Data Processing Training Program of the University of Michigan, Ann Arbor, Michigan: Malloy Lithographing Inc.
- JOHNSON, P.E., MOLLER, J.H. and Bass, G.M. (1975) Analysis of expert diagnosis of a computer simulation of congenital heart disease. J. of Med. Educ., 50, 5, 466-470.
- JOLLY, B.C. (1976) Assessment: techniques of assessment 2. Multiple choice questions. Data Sheet No. 3. Information, 6, 15-16.
- JOLLY, B.C. and GALE, JANET (1976) Statistical significance: its appropriate use. Med. Educ., 10, 5, 410-411.
- JONES, D.P., CHIPPER, L.M. and HOLZWORTH, R.J. (1978) Effects of amount of information on decision strategies. J. of Gen. Psychol., 98, 281-294.
- JUUL, D.H., NOE, M.J. and NERENBERG, R.L. (1979) A factor analytic study of branching patient management problems. Med. Educ., 13, 3, 199-203.
- KAGAN, N. and KRATHOHL, D.R. (1967) Studies in Human Interaction: Interpersonal Process Recall Stimulated by Videotape. Final Report, Project No. 5-0830 U.S. Department of Health, Education and Welfare.
- KAGAN, N. (1977) IPR-media in clinical and human interaction supervision. Prepublication discussion paper.
- KEGEL-FLOM, P. (1975) Predicting supervisor, peer and self-ratings of intern performance. J. of Med. Educ., 50, 812-815.

- KLECKA, W.R. (1975) Discriminant analysis. In N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner and D.H. Bent. Statistical Package for the Social Sciences. 2nd edition. New York: McGraw-Hill Inc.
- KLEIN, S.P., FREDERIKSON, N. and EVANS, F.R. (1969) Anxiety and learning to formulate hypotheses. J. of Educ. Psychol., 60, 6, 465-475.
- KLEINMUNTZ, B. (1968) Formal Representation of Human Judgement. New York: Wiley.
- KNAFL, K. and BURKETT, G. (1975) Professional socialisation in a surgical specialty: Acquiring medical judgement. Soc. Sci. and Med., 9, 397-404.
- KNAGGS, S.J., BARNES, A.J. and MACLEAN, C.B. (1974) The decision tree for teaching management of uncertainty. J. of Med. Educ., 49, 12, 1184-1186.
- KNILL-JONES, R.P., STERN, R.B., GIRMES, D.H., MAXWELL, J.D., THOMPSON, R.P.H. and WILLIAMS, R. (1973) Use of sequential Bayesian model in diagnosis of jaundice by computer. Brit. Med. J., 1, 530-533.
- KNILL-JONES, R.P. (1977) Clinical decision-making (2) Diagnostic and prognostic inference. Health Bull., 35, 4, 213-222.
- KULIKOWSKI, C.A. (1970) Pattern recognition approach to medical diagnosis. IEEE Trans. on Syst. Sci. and Cybern., 6, 3, 173-178.
- LAVELLE, S.M. (1977a) A course in objective methods for clinical students. Paper read at the Annual Scientific Meeting of the Association for the Study of Medical Education, Leeds.
- LAVELLE, S.M. (1977b) Information processing by medical students. Paper presented at the Conference of the Association for Medical Education in Europe, Uppsala.
- LAVELLE, S.M. (1978) Information processing by medical students. Med. Educ., 12, 5, Supplement, 108-109.
- LAWLESS, C.J. (1979) Information processing: A model for educational technology? Paper presented at the International APLET Conference, Sheffield City Polytechnic.
- LEAPER, D.J., GILL, P.W., STANILAND, J.R., HORROCKS, J.C., and de DOMBAL, F.T. (1973) Clinical diagnostic process: An analysis. Brit. Med. J., 3, 569-574.
- LEARNING RESOURCES DESIGN PROJECT (Programme for Educational Development) (1975) Guide for the Use of Problem Based Learning Units on the Nervous System No. 1. McMaster University, Faculty of Health Sciences.
- LEDLEY, R.S. and LUSTED, L.B. (1959) Reasoning foundations of medical diagnosis. Science, 130, 9-21.
- IFINNOX, B. (1967) Marking multiple choice examinations. Brit. J. of Med. Educ., 1, 3, 203-211.
- LEWY, A. (1973) Discrimination among individuals vs. discrimination among groups. J. of Educ. Meas., 10, 1, 19-24.
- LINDBERG, D.A.B. (1968) The Computer and Medical Care. Springfield, Illinois: Charles C. Thomas.
- LISACK, T. and FU, K.S. (1976) Parametric feature extraction through error minimization applied to medical diagnosis. IEEE Trans. on Syst., Man and Cybern., 6, 9, 605-611.
- LLOYD, G., TYN PUGH, E. and MCINTYRE, N. (1976) The Problem Orientated Medical Record and its Educational Implications. ASME Medical Education Booklet No. 6. Dundee: ASME.

- LORD, F.M. (1975) Formula scoring and number-right scoring. J. of Educ. Meas., 12, 1, 7-11.
- LOVELL, K. (1961) A follow-up study of Inhelder and Piaget's "The Growth of Logical Thinking". Brit. J. of Psychol., 52, 2, 143-153.
- LOVELL, K. (1974) Intellectual aspects of adolescence. In M. Kellmer Pringle and V. Varma (Eds) Advances in Educational Psychology Part 2. London: University of London Press Ltd.
- LUCHINS, A.S. and LUCHINS, E.H. (1950) New experimental attempts at presenting mechanization in problem solving. In P.C. Wason and P.N. Johnson-Laird (Eds) Thinking and Reasoning. Harmondsworth: Penguin Books Ltd.
- LUNZER, E.A. (1965) Problems of formal reasoning. Monogr. of the Soc. for Res. in Child Dev., 100, 19-46.
- LUNZER, E.A. (1968) Formal reasoning. In E.A. Lunzer and J.F. Morris (Eds) Development in Human Learning. London: Staples.
- IUSTED, L.B. and STAHL, J.R. (1964) Conceptual models of diagnosis. In J.A. Jacquez (Ed) The Diagnostic Process. Proceedings of a conference sponsored by the Biochemical Data Processing Training Program of the University of Michigan. Ann Arbor, Michigan: Malloy Lithographing Inc.
- MACE, E. (1976) Rolling remake: An alternative approach to course design at the Open University. Prog. Learning and Educ. Tech., 13, 4, 25-30.
- MADDISON, D.C. (1978) What's wrong with medical education? Med. Educ., 12, 3, Supplement, 28-33.
- MAGUIRE, G.P., CLARKE, D. and JOLLY, B. (1977) An experimental comparison of three courses in history-taking skills for medical students. Med. Educ., 11, 3, 175-182.
- MAIER, N.R.F. (1931) Reasoning in humans II. The solution of a problem and its appearance in consciousness. J. of Comp. Psychol., 12, 181-194.
- MANIS, M. (1968) Cognitive Processes. Belmont, California: Brooks/Cole Publishing Company.
- MARSHALL, V.R. and LUDBROOK, J. (1972) The relative importance of patient and examiner variability in a test of clinical skills. Brit. J. of Med. Educ., 6, 3, 364-368.
- MARTON, F. (1975) What does it take to learn? In N.J. Entwistle and D.J. Hounsell (Eds) How Students Learn. Lancaster: Institute for Post-Compulsory Education.
- MARTON, F. and SALJO, R. (1976a) On qualitative differences in learning. 1 - Outcome and process. Brit. J. of Educ. Psychol., 46, 1, 4-11.
- MARTON, F. and SALJO, R. (1976b) On qualitative differences in learning. 2 - Outcome as a function of the learner's conception of the task. Brit. J. of Educ. Psychol., 46, 2, 115-127.
- MAYOU, R. (1978) Psychiatric decision-making by medical students. Brit. J. of Psychiat., 132, 191-194.
- MCCARTHY, W.H. (1966) An assessment of the influence of cuing items in objective examinations. J. of Med. Educ., 41, 262-266.
- MCCARTHY, W.H. and GONELLA, S.S. (1967) The simulated patient management problem: A technique for evaluating and teaching clinical competence. Brit. J. of Med. Educ., 1, 348-352.

- McGUIRE, C.H. and SOLOMON, L.M. (1971) Clinical Simulations. New York: Appleton, Century and Crofts.
- McKEACHIE, W.J. (1963) Research on teaching at the college and university level. In N.L. Gage (Ed) Handbook of Research on Teaching. Chicago: Rand McNally and Co.
- McNEMAR, Q. (1962) Psychological Statistics. New York: John Wiley and Sons Inc.
- McWHINNEY, I.R. (1972) Problem solving and decision making in primary medical practice. Proc. of the Roy. Soc. of Med., 65, 934-938.
- MEADOR, C.K. (1969) Nondisease: a problem of overdiagnosis. Diagnostica, 14, 10-11.
- MEALS, R.A. (1973) Teaching clinical judgement. Brit. J. of Med. Educ., 7, 2, 100-102.
- MEDAWAR, P.B. (1969) Induction and Intuition in Scientific Thought. London: Methuen and Co. Ltd.
- MEEHL, P.E. (1954) Clinical Versus Statistical Prediction. Minneapolis: University of Minnesota Press.
- MEEHL, P.E. (1960) The cognitive activity of the clinician. Amer. Psychologist, 15, 19-27.
- MILLER, G.A., GALANTER, E. and PRIBRAM, K.H. (1960) The formation of plans. In P.C. Wason and P.N. Johnson-Laird (Eds) Thinking and Reasoning. Harmondsworth: Penguin Books Ltd.
- MILLER, G.A. (1962) Psychology: The Science of Mental Life. Harmondsworth: Penguin Books Ltd.
- MILLER, G.E., GRASER, H.P., ABRAHAMSON, S., HARNACK, R.S., COHEN, I.S., and LAND, A. (1961) Teaching and Learning in Medical School. Cambridge, Mass.: Harvard University Press.
- MILLER, G.E. (1978) The contribution of research in the learning process. Med. Educ., 12, 3, Supplement, 28-33.
- MORTON, J. (1967) A singular lack of incidental learning. Nature, 215, 203-204.
- MURRAY, T.S., CUPPLES, R.W., BARBER, J.H., DUNN, W.R., COTT, D.B. and HAINAY, D.R. (1977) Teaching decision making to medical undergraduates by computer-assisted learning. Med. Educ., 11, 4, 262-264.
- MYLLER-CROOK, C. (1974) Behavioural science for medical students. Bull. of the Brit. Psychol. Soc., 27, 403-406.
- NEISSER, U. (1967) Cognitive Psychology. New York: Appleton - Century - Crofts.
- NEBLE, D.I. (1975) Evaluation of Clinical Competence. Study Leave Report. The University of Adelaide, Department of Medicine.
- NEWELL, A. and SIMON, H.A. (1972) Human Problem Solving. Englewood Cliffs, New Jersey: Prentice-Hall Inc.
- NEUFELD, V.R., NORMAN, G.R., FEIGHTNER, J.W. and BARROWS, H.S. (1976) Clinical methods of medical students. 2. A cross-sectional analysis. Unpublished report.
- NISBETT, R.E. and WILSON, T.D. (1977) Telling more than we know. Verbal reports on mental processes. Psychol. Rev., 84, 3, 231-259.
- NIXON, M. (1973) Classification task performance and Piagetian theory. Australian Psychologist, 8, 2, 85-99.

- NORMAN, G.R., BARROWS, H.S., NEUFELD, V.R. and FEIGHTNER, J.W. (1977) A data base for analysis of the process of clinical problem solving. Proceedings, Medinfo '77 Amsterdam: Shires and Wolf, North Holland.
- NORMAN, G.R., BARROWS, H.S., NEUFELD, V.R. and FEIGHTNER, J.W. (undated) A data base for analysis of the process of clinical problem solving. Unpublished report.
- NYSTEDT, L. (1972) Judgements in quasi-clinical situations: Accuracy and process. Reports from the Psychological Laboratories, The University of Stockholm. Supplement series. Supplement 12.
- NYSTEDT, L. and MAGNUSSON, D. (1972) Predictive efficiency as a function of amount of information. Multivar. Beh. Res., 7, 441-450.
- NYSTEDT, L. (1974) Consensus among judges as a function of amount of information. Educ. and Psychol. Meas., 34, 91-101.
- NYSTEDT, L. and MAGNUSSON, D. (1975) Integration of information in a clinical judgement task: an empirical comparison of six models. Percept. and Motor Skills, 40, 343-356.
- OHNMACHT, F.W. (1966) Effects of field independence and dogmatism on reversal and non-reversal shifts in concept formation. Percept. and Motor Skills, 22, 491-497.
- OLSON, I.A. (1976) The organisation of an integrated basic medical science curriculum at a new medical school, Nottingham. Med. Educ., 10, 5, 362-368.
- OPPENHEIM, A.N. (1966) Questionnaire Design and Attitude Measurement. London: Heinemann.
- OSGOOD, C.E. (1953) Method and Theory in Experimental Psychology. New York: Oxford University Press.
- OSKAMP, S. (1965) Overconfidence in case study judgements. J. of Consult. Psychol., 29, 261-265.
- PALVA, I.P. and KORHONEN, V. (1973) Confidence testing as an improvement of multiple choice examinations. Brit. J. of Med. Educ., 7, 3, 179-181.
- PARRY, K.M. (1978) Changing medical education in western Europe. Introduction. Med. Educ., 12, 3, Supplement, 2-3.
- PAUKER, S.G., GORRY, G.A., KASSIRER, J.P. and SCHWARZ, W.B. (1976) Towards the simulation of clinical cognition. Taking a present illness by computer. Amer. J. of Med., 60, 981-996.
- PAXTON, R. (1976) Note. Some criteria for choosing between explanations in psychology. Bull. of the Brit. Psychol. Soc., 29, 396-399.
- PEEL, E.A. (1967) The Psychological Basis of Education. Edinburgh: Oliver and Boyd Ltd.
- PEEL, E.A. (1968) Conceptual learning and explainer thinking. In E.A. Lunzer and J.F. Morris (Eds) Development in Human Learning. Volume 2. London: Staples Press.
- PEEL, E.A. (1971) The Nature of Adolescent Judgment. London: Staples Press.
- PETERSON, P.L. and CLARK, C.M. (1973) Teachers' reports of their cognitive processes during teaching. Amer. Educ. Res. J., 15, 4, 555-565.
- PIAGET, J. (1968) Structuralism. London: Routledge and Kegan Paul.
- PIAGET, J. (1970) Science of Education and the Psychology of the Child. London: Longman.

- POOL, I. de S. (1959) (Ed) Trends in Content Analysis. Urbana: University of Illinois Press.
- POSTMAN, L. and TOLMAN, E.C. (1959) Brunswik's probabilistic functionalism. In S. Koch (Ed) Psychology: A Study of a Science. Study 1. Conceptual and Systematic. Volume 1. Sensory, Perceptual and Physiological Formulations. New York: McGraw-Hill.
- POUND, A. (1978) In defence of eclecticism. A clinical psychologist's view of psychological theories. Bull. of the Brit. Psychol. Soc., 31, 360-361.
- RADFORD, J. (1974) Reflections on introspection. Amer. Psychologist, 29, 245-250.
- REITMAN, M.R. (1964) Heuristic decision procedures, open constraints and the structure of ill-defined problems. In M.W. Shelly and G.L. Bryan (Eds). Human Judgements and Optimality. New York: Wiley and Sons Inc.
- RICHARDSON, H.P. (1977) Non-metric multi-dimensional scaling as an aid in the analysis of open-ended interview data. Paper read at the Annual Conference of the British Psychological Society, Exeter.
- RIEGER, K.F. (1973) Dialectic operations. The final period of cognitive development. Human Dev., 17, 386-391.
- RIMOLDI, H.J.A. (1955) A technique for the study of problem solving. Educ. and Psychol. Meas., 15, 450-461.
- RIMOLDI, H.J.A. (1961) The test of diagnostic skills. J. of Med. Educ., 36, 73-79.
- RIMOLDI, H.J.A. (1964) Testing and analysis of diagnostic skills. In J.A. Jacquez (Ed) The Diagnostic Process. Proceedings of a conference sponsored by the Biochemical Data Processing Training Program of the University of Michigan. Ann Arbor, Michigan: Malloy Lithographing Inc.
- ROTHMAN, A.J. (1969) Confidence testing: an extension of multiple choice testing. Brit. J. of Med. Educ., 3, 4, 237-239.
- ROWLEY, G.L. and TRAUB, R.E. (1977) Formula scoring, number-right scoring and test-taking strategy. J. of Educ. Meas., 14, 1, 15-21.
- ROYAL COLLEGE OF GENERAL PRACTITIONERS (1972) The Future General Practitioner: Learning and Teaching. London: British Medical Journal.
- ROYAL FREE HOSPITAL SCHOOL OF MEDICINE (1978) The New Clinical Curriculum (Yellow Booklet) Internal publication.
- RUFF, G.E. and MECHANICK, P.G. (1975) Teaching psychological diagnosis and management. Brit. J. of Med. Educ., 9, 3, 152-157.
- SALJO, R. (1978) Learning about learning. Paper presented at the Fourth International Conference on Higher Education, University of Lancaster.
- SANDERSON, P.H. (1973) The 'don't know' option in MCQ examinations. Brit. J. of Med. Educ., 7, 1, 25-29.
- SARBIN, T.R., TAFT, R. and BATLEY, P.E. (1960) Clinical Inference and Cognitive Theory. New York: Holt, Rinehart and Winston Inc.
- SATYER, J. (1966) Measurement and prediction, clinical and statistical. Psychol. Bull., 66, 3, 178-200.
- SCHADDING, J.G. (1967) Diagnosis: The clinician and the computer. The Lancet, 1, 877-882.
- SCHNEIDERMAN, H. and MULLER, R.L. (1972) The diagnosis game. J. of the Amer. Med. Assoc., 219, 333-335.

- SCHOOLS' COUNCIL (1965) The certificate of secondary education: Experimental examinations - Mathematics 2. Examinations Bulletin 7. London: Schools' Council.
- SCHRODER, H.M., DRIVER, M.J. and STREUFERT, S. (1967) Human Information Processing. New York: Holt, Rinehart and Winston.
- SCHWARTZ, W.B., GORRY, G.A., KASSIRER, J.P. and ESSIG, A. (1973) Decision analysis and clinical judgement. Amer. J. of Med., 55, 459-472.
- CHAPSON, S.M. (1977) Hypothesis testing and cognitive style in children. J. of Educ. Psychol., 69, 4, 452-463.
- SHAVE' SON, R.J. and STANTON, G.C. (1975) Construct validation: Methodology and application to three measures of cognitive structure. J. of Educ. Meas., 12, 2, 67-85.
- SHERWOOD, T. (1978) Science in radiology. The Lancet, 1, 594-595.
- SHULMAN, L.S. (1972) Psychological controversies in teaching. In M.L. Silberman, J.S. Allender and J.M. Yanoff (Eds) The Psychology of Open Teaching and Learning. An Inquiry Approach. Boston: Little, Brown and Company.
- SHULMAN, L.S. and ELSTEIN, A.S. (1975) Studies of problem solving, judgement and decision making: Implications for educational research. Rev. of Res. in Educ., 3, 3-42.
- SIBLEY, J. (1978) Faculty of Health Sciences, McMaster University, Canada - the 1977 perspective. Med. Educ., 12, 5, Supplement, 15-18.
- SIEGEL, L., SIEGEL, L.C., CAPRETTA, P.J., JONES, R.I. and BERKOWITZ, H. (1963) Students' thoughts during class: A criterion for educational research. J. of Educ. Psychol., 54, 1, 45-51.
- STIEGER, S. (1956) Nonparametric Statistics for the Behavioural Sciences. New York: McGraw-Hill Book Company Inc.
- SIMBORG, D.W., STARFIELD, B.H., HORN, S.D. and YOURTREE, S.A. (1976) Information factors affecting problem follow-up in ambulatory care. Med. Care, 14, 10, 848-856.
- SIMPSON, M.A. (1971) Medical Education. A Critical Approach. London: Butterworths.
- SMART, G.A. (1978) The nature of graduate training. Med. Educ., 12, 3, Supplement, 5.
- SMEDSLUND, J. (1963) The concept of correlation in adults. Scand. J. of Psychol., 4, 165-173.
- PLEDIE, S.M., HOUTZ, J.C., RINGENBACH, S. and FELDHOUSEN, J.F. (1973) Abilities measured by the Purdue Elementary Problem-solving Inventory. Psychol. Rep., 33, 959-963.
- SPIELMAN, E.L. (1965) The administration of a co-ordinated teaching curriculum. J. of Med. Educ., 40, 11, 1049-1057.
- SPRAFKA, S.A. and ELSTEIN, A. (1974) What do physicians do? An analysis of diagnostic reasoning. Paper presented at the OMERAD Symposium on Medical Judgement, Michigan State University, East Lansing.
- TENHOUSE, L. (1975) An Introduction to Curriculum Research and Development. London: Heinemann.
- TONES, E. (1966) An Introduction to Educational Psychology. London: Methuen and Co. Ltd.

- TREUFERT, C.C. (1973) Effects of information relevance on decision making in complex environments. Memory and Cog., 1, 3, 224-228.
- TROM, G. and WALTON, H.J. (Eds) (1978) Innovations in medical education. Med. Educ., 12, 5, Supplement.
- TURGEON, D.A. (1979) Videotapes in psychiatry: Their use in teaching observation techniques. Med. Educ., 13, 3, 204-208.
- SVENSSON, L. (1977) Study skill and learning. Brit. J. of Educ. Psychol., 47, 3, 233-243.
- TATSUOKA, M.M. (1971) Multivariate Analysis: Techniques for Educational and Psychological Research. New York: John Wiley and Sons Inc.
- TAYLOR, P.J., HARASYM, P.H. and LAURENSEN, R.D. (1978) Introducing first-year medical students to early diagnostic hypotheses. J. of Med. Educ., 53, 402-409.
- TAYLOR, T.R. (1970a) Computer-guided diagnosis. J. of the Roy. Coll. of Physicians of London, 4, 3, 188-195.
- TAYLOR, T.R. (1970b) Computers in medicine. Science J., 6, 81-86.
- TAYLOR, T.R. (1971) Models of the diagnostic process. Unpublished PhD. thesis. Glasgow University.
- TAYLOR, T.R., AITCHISON, J. and MCGIRR, E.M. (1971) Doctors as decision-makers. A computer-assisted study of diagnosis as a cognitive skill. Brit. Med. J., 3, 35-40.
- TAYLOR, T.R. (1972) Computer-assisted instruction in medical education. Prog. Learning and Educ. Tech., 9, 5, 272-282.
- THOMSON, R. (1959) The Psychology of Thinking. Harmondsworth: Penguin Books Ltd.
- THWATTES, J.G. (1954) Into General Practice. London: Heinemann.
- THYNE, J.M. (1966) The Psychology of Learning and Techniques of Teaching. London: University of London Press.
- TURNEY, B.L. and ROBB, G.P. (1971) Research in Education: An Introduction. Hinsdale, Illinois: The Dryden Press Inc.
- VALENTINE, E.R. (1978) Perchings and flights: The contribution of introspection to the study of thinking. In A. Burton and J. Radford (Eds) Thinking in Perspective. London: Methuen.
- VAUGHAN, M. and MARKS, J.N. (1976) Teaching interviewing skills to medical students: A comparison of two methods. Med. Educ., 10, 3, 170-175.
- VINACKE, W.E. (1952) The Psychology of Thinking. New York: McGraw-Hill.
- WASON, P.C. (1968) On the failure to eliminate hypotheses. A second look. In P.C. Wason and P.N. Johnson-Laird (Eds) Thinking and Reasoning. Harmondsworth: Penguin Books Ltd.
- WED, L.L. (1968) Medical records that guide and teach. The New Eng. J. of Med., 278, 11, 593-600 and 652-657.
- WED, L.L. (1969) Medical Records, Medical Education and Patient Care: The Problem-Oriented Record as a Basic Tool. Cleveland, Ohio; Press of Case Western Reserve University.
- WERTHEIMER, M. (1945) Productive Thinking. New York: Harper and Brothers.
- WESMAN, A.G. (1971) Writing the test item. In R.L. Thorndike (Ed) Educational Measurement. Second Edition. Washington D.C.: American Council on Education.

- WHITBY, L.G. (1977) Marking systems for multiple choice examinations. Med. Educ., 11, 3, 216-220.
- WIGGINS, N. and HOFFMAN, P.J. (1968) Three models of clinical judgement. J. of Abnorm. Psychol., 73, 70-77.
- WINER, B.J. (1972) Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company Inc.
- WITKIN, H.A., DYK, R.B., FATERSON, H.F., GOODENOUGH, D.R., and KARP, S.A. (1974) Psychological Differentiation. New York: John Wiley and Sons Inc.
- WOMERSLEY, J., STENHOUSE, G. and DUNN, W.R. (1974) Use of a response system in university lectures: aid to structuring of information. Brit. J. of Med. Educ., 8, 3, 192-196.
- WOOD, D., BRUNER, J.S. and ROSS, G. (1976) The role of tutoring in problem solving. J. of Child Psychol. and Psychiat., 17, 2, 89-100.
- WOOD, R. (1977) Multiple choice. A state of the art report. Eval. in Educ., 1, 191-280. Oxford, England: Pergamon Press.
- WYN PUGH, E., LLOYD, G.J. and MCINTYRE, N. (1975) Relevance of educational objectives for medical education. Brit. Med. J., 3, 688-691.
- YONKE, A.M. (1979) The art and science of clinical teaching. Med. Educ., 13, 2, 86-90.
- YOUNISS, J. (1974) Operations and everyday thinking. A commentary on "Dialectical Operations". Human Dev., 17, 386-391.